

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 28, 2003, 10:33:50 ; Search time 1527.67 Seconds

(without alignments)
285.757 Million cell updates/sec

Title: SEQ1-65to79
Perfect score: 15
Sequence: 1 SURSAHLAQSIILSC 15

Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 8

Total number of hits satisfying chosen parameters: 107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODE=frame+pn.model -DEV=xlh
-Q/cgnt2.1/USPTO.spool/BORN62/Runat_23012003.130141.7873/app.query.fasta.1.597
-DB=GenEmbl -OFMT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human4.0.cdi -LIST=1000
-DOCALLIGN=200 -THR=SCORE=quality -THR_MIN=8 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=BORN62.GCGN.1.1.2425 @runat.23012003.130141.7873 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEOUTER -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEX=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vl.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vl.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pin.*
35: em.htg_rtd.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em_sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	723	1	EMDLTA	M57244 Plasmid EMP
2	15	100.0	777	1	AF242417	AF242417 Escherich
3	15	100.0	777	1	P307LTPA	M35581 Plasmid p30
4	15	100.0	777	6	A04913	A04913 E.coli LTA
5	15	100.0	777	6	A04915	A04915 Artificial
6	15	100.0	866	1	ENSLTTPA	M15362 Plasmid ENT
7	15	100.0	898	1	ENRLTTPA	K01995 E.coli toxa
8	15	100.0	934	1	ECOTOXA	E03421 DNA encodin
9	15	100.0	1148	6	E03421	S60731 heat-labile
10	15	100.0	1275	1	S60731	AB011677 Escherich
11	15	100.0	1434	1	AB011677	AB011677 A subunit O
12	10	66.7	711	6	AR1818597	AR1818597 Sequence
13	10	66.7	711	6	AR1818597	V00275 E. coli gen
14	10	66.7	868	1	ECEHTA	E00613 DNA encodin
15	9	60.0	777	6	E00613	AC114463 Rattus no
16	9	60.0	155416	2	AC114463	AC120962 Rattus no
17	9	60.0	174342	2	AC120962	BC021349 Mus muscu
18	8	53.3	1877	10	BC021349	BC004766 Mus muscu
19	8	53.3	3014	10	BC004766	U10336 Helobdella
20	8	53.3	4011	3	HSU10336	AE013887 Yersinia
21	8	53.3	9968	1	AE013887	AC004462 Homo sap1
22	8	53.3	36021	9	AC004462	AC130257 Rattus no
23	8	53.3	51644	2	AC130257	AC111766 Rattus no
24	8	53.3	72519	2	AC111766	AC115213 Rattus no
25	8	53.3	81938	2	AC115213	AC120099 Rattus no
26	8	53.3	83292	2	AC120099	AC121484 Rattus no
27	8	53.3	85022	2	AC121484	AC105557 Rattus no
28	8	53.3	93766	2	AC105557	AC111610 Rattus no
29	8	53.3	99537	2	AC111610	AC113657 Rattus no
30	8	53.3	100786	2	AC113657	AC107589 Rattus no
31	8	53.3	105756	2	AC107589	AC094679 Rattus no
32	8	53.3	107637	2	AC094679	L77570 Homo sapien
33	8	53.3	108400	9	HUMDGCRCEN	AC128996 Rattus no
34	8	53.3	108551	2	AC128996	AC026736 Homo sap1
35	8	53.3	108661	9	AC026736	AC109707 Rattus no
36	8	53.3	109396	2	AC109707	AC128969 Rattus no
37	8	53.3	114789	2	AC128969	AC131517 Rattus no
38	8	53.3	118681	2	AC131517	AC157764 Human DNA
39	8	53.3	119785	2	AC157764	AC108338 Rattus no
40	8	53.3	121080	2	AC108338	AC120314 Rattus no
41	8	53.3	122517	2	AC091350	AC091350 Rattus no
42	8	53.3	124579	2	AC103101	AC103101 Rattus no
43	8	53.3	126809	2	AC124138	AC124138 Rattus no
44	8	53.3	137910	9	AL157764	AL157764 Human DNA
45	8	53.3	147108	2	AC108338	AC115134 Rattus no
46	8	53.3	148879	2	AC115134	AL354740 Human DNA
47	8	53.3	151828	2	AL354740	AL355871 Human DNA
48	8	53.3	152191	2	AL355871	AC004846 Homo sap1
49	8	53.3	154588	9	AC004846	
50	8	53.3	154616	9	AC004846	

C	51	8	53.3	154678	2	AC128512	AC128512 Rattus no
C	52	8	53.3	154737	2	AC129370	AC129370 Rattus no
C	53	8	53.3	155284	2	AC109896	AC109896 Rattus no
C	54	8	53.3	155542	2	AC130769	AC130769 Rattus no
C	55	8	53.3	155846	2	AC103239	AC103239 Rattus no
C	56	8	53.3	155904	2	AC015953	AC015953 Homo sapi
C	57	8	53.3	157328	2	AC096203	AC096203 Rattus no
C	58	8	53.3	158371	2	AC011155	AC011155 Homo sapi
C	59	8	53.3	158478	2	AC101114	AC101114 Rattus no
C	60	8	53.3	160770	2	AC106148	AC106148 Rattus no
C	61	8	53.3	161273	9	HS769813	HS769813 Human DNA
C	62	8	53.3	161396	2	AC110971	AC110971 Rattus no
C	63	8	53.3	161489	2	AC118332	AC118332 Rattus no
C	64	8	53.3	164812	2	AC120628	AC120628 Rattus no
C	65	8	53.3	165581	2	AC111952	AC111952 Rattus no
C	66	8	53.3	166772	2	AC098691	AC098691 Homo sapi
C	67	8	53.3	167653	2	AC108983	AC108983 Rattus no
C	68	8	53.3	168813	2	AC103484	AC103484 Rattus no
C	69	8	53.3	170678	2	AC111666	AC111666 Rattus no
C	70	8	53.3	174666	2	AC111212	AC111212 Rattus no
C	71	8	53.3	175414	2	AC102971	AC102971 Rattus no
C	72	8	53.3	175671	2	AC010274	AC010274 Homo sapi
C	73	8	53.3	175948	2	AC095485	AC095485 Rattus no
C	74	8	53.3	176233	2	AC024449	AC024449 Homo sapi
C	75	8	53.3	176234	2	AC016383	AC016383 Homo sapi
C	76	8	53.3	178809	2	AC098062	AC098062 Rattus no
C	77	8	53.3	179281	2	AP001934	AP001934 Homo sapi
C	78	8	53.3	180353	2	AC121185	AC121185 Rattus no
C	79	8	53.3	180401	2	AC096933	AC096933 Rattus no
C	80	8	53.3	181904	2	AC027623	AC027623 Homo sapi
C	81	8	53.3	182926	2	AC120476	AC120476 Rattus no
C	82	8	53.3	184140	2	AC121479	AC121479 Rattus no
C	83	8	53.3	186973	2	AC124932	AC124932 Rattus no
C	84	8	53.3	187274	2	AL391096	AL391096 Homo sapi
C	85	8	53.3	188133	2	AC115147	AC115147 Rattus no
C	86	8	53.3	188762	2	AC097821	AC097821 Rattus no
C	87	8	53.3	188859	2	AC118091	AC118091 Rattus no
C	88	8	53.3	189877	2	AC128729	AC128729 Rattus no
C	89	8	53.3	191024	2	AC126651	AC126651 Rattus no
C	90	8	53.3	196626	2	AC125668	AC125668 Rattus no
C	91	8	53.3	199050	1	AF141417	AF141417 Yersinia
C	92	8	53.3	199852	2	AC118406	AC118406 Rattus no
C	93	8	53.3	200832	2	AC130970	AC130970 Rattus no
C	94	8	53.3	202842	10	AL607108	AL607108 Mouse DNA
C	95	8	53.3	203046	2	AC006342	AC006342 Homo sapi
C	96	8	53.3	203875	2	AC128558	AC128558 Rattus no
C	97	8	53.3	204233	2	AC095457	AC095457 Rattus no
C	98	8	53.3	204302	2	AC068650	AC068650 Mus muscu
C	99	8	53.3	205571	2	AC118497	AC118497 Rattus no
C	100	8	53.3	210000	2	AC007506	AC007506 Homo sapi
C	101	8	53.3	211208	2	AC094931	AC094931 Rattus no
C	102	8	53.3	213351	2	AC111235	AC111235 Rattus no
C	103	8	53.3	216387	9	DJ534K4	AF109907 Homo sapi
C	104	8	53.3	219771	2	AC097232	AC097232 Rattus no
C	105	8	53.3	228563	2	AC090219	AC090219 Homo sapi
C	106	8	53.3	234787	2	AC073780	AC073780 Mus muscu
C	107	8	53.3	247509	2	AC127411	AC127411 Mus muscu

ALIGNMENTS

RESULT 1
EMBLTA 723 bp DNA linear BCT 23-APR-1996
LOCUS Plasmid EMD 299 (from E.coli) heat-labile enterotoxin subunit A
DEFINITION (17-A) gene, 3' end.
ACCESSION M57244.1 GI:531190
VERSION M57244.1 GI:531190
KEYWORDS enterotoxin; heat-labile enterotoxin.
SOURCE Plasmid EMD 299 DNA.
ORGANISM plasmids.
REFERENCE 1 (bases 1 to 723)

AUTHORS Tsuji,T., Inoue,T., Miyama,A., Okamoto,K., Honda,T. and Miwatani,T.
TITLE A single amino acid substitution in the A subunit of Escherichia coli enterotoxin results in a loss of its toxic activity
JOURNAL J Biol. Chem. 265 (36), 22520-22525 (1990)
MEDLINE 91093102
COMMENT On Aug 20, 1994 this sequence version replaced gi:148531.
PUBMED 2265142
FEATURES
source location/Qualifiers
1..723
/organism="Plasmid EMD 299"
/specific_host="Escherichia coli"
/db_xref="taxon:2464"
/plasmid="Plasmid EMD 299"
1..723
/gene="17-A"
/gene="17-A"
1..723
/gene="17-A"
/codon_start=1
/transl_table=11
/product="heat-labile enterotoxin subunit A"
/protein_id="AAB59161.1"
/db_xref="GI:148532"
/translation="MGDRLYRADSRPPDEIKRSGGLMPRGHNEYRDRCTONINLYDH
AREQTGFVYRDPDGYVSTLSLSAHLAQSTLSGYSFYIVATAPNMNVNDVLG
VSPHYEQKVSALGIPYSQIYGVYRVNFVGIIDRLHGRNREYDRYRNINIAED
GYRLAGFPDHOAMREEPWIHHAPGCGMSNTITIGDNCNEQTMLSTIYLREYQSKV
KRGIQSDYQSEVDIYNRIKDEL"
459
old_sequence
/gene="17-A"
/note="A"
/citation={1}
/replace=""
BASE COUNT 238 a 128 c 160 g 197 t
ORIGIN
Alignment Scores:
Pred. No.: 5.05e-07 Length: 723
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
SEQ1-65T079 (1-15) x EMDLTA (1-723)
QY 1 SerleuargserAlahisteAlahylgInserileuSercly 15
|||||
Db 193 AGTTGAGAGTCTCATTAGCAGCAGCTATATACAGA 237
RESULT 2
AF242417 777 bp DNA linear BCT 01-JAN-2002
LOCUS Escherichia coli heat-labile enterotoxin subunit A precursor, gene,
DEFINITION complete cds.
ACCESSION AF242417
VERSION AF242417.1 GI:18026885
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 777)
AUTHORS Chang,G.N. and Ho,K.C.
TITLE Heat-labile enterotoxin subunit A gene of Escherichia coli
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 777)
AUTHORS Chang,G.N. and Ho,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2000) Botany, National Taiwan University, 1
Roosevelt Road, Sec. 4, Taipei, Taiwan 106, Republic of China
FEATURES
source location/Qualifiers
1..777
/organism="Escherichia coli"

```

CDS
/strain="CH-C2"
/serotype="O-136"
/specific_host="pig"
/db_xref="taxon:562"
/country="Taiwan"
/notes="Isolated from hemolytic cells"
1..777
/contig="ltpa"
/codon_start=1
/transl_table=11
/product="heat-labile enterotoxin subunit A precursor"
/protein_id="AAL5671.1"
/db_xref="GI:18026886"
/translation="MKNTFFIFILLASPLVANGDRLYRADSRPDEIKRSGGLMPRG
HNEYFDRTGMNINLYDHARGTQTFVRVDDGYSTSLRSASHLAGOSILSGSYTY
IYVATAPMFMFNVNDLVGYSPHPYEQEVSALGGLPYQIYGMVYVNFVDERLARN
REYRDRIYRNINLAPAEDEGRLAGPPDQANREEPWIHHAQGGCNSSRTTGDTCN
EETQNLSTIYLEYOSKVRQIFSDYQSEVDIYNIRDEL"
1..54
sig_peptide
55..630
mat_peptide
/product="heat-labile enterotoxin subunit A1"
637..774
mat_peptide
/product="heat-labile enterotoxin subunit A2"
136 c 164 g 222 t

BASE COUNT 255 a 136 c 164 g 222 t
ORIGIN

Alignment Scores:
Pred. No.: 5.38e-07 Length: 777
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

SEQ1-65TO79 (1-15) x AF242417 (1-777)
QY 1 SerleuArgSerAlaHisLeuAlaGlyGlnSerIleuSergly 15
|||||
Db 247 AGTTTGAGAGTGTCTCATTAGCAGCAGCTATATTATCAGCA 291

RESULT 3
LOCUS P307LTA 777 bp DNA linear BCT 24-APR-1996
DEFINITION Plasmid P307 (from E. coli) heat-labile enterotoxin subunit A (lta)
ACCESSION M35381.1 GI:150458
VERSION M35381.1 GI:150458
KEYWORDS enterotoxin; lta gene.
SOURCE Plasmid P307 (clone: PAT153.) DNA.
ORGANISM plasmids.
REFERENCE 1 (bases 1 to 777)
AUTHORS Dykes,C.W., Halliday,I.J., Hobden,A.N., Read,M.J. and Harford,S.
TITLE A comparison of the nucleotide sequence of the A subunit of heat-
labile enterotoxin and cholera toxin
JOURNAL FEMS Microbiol. Lett. 26, 171-174 (1985)
FEATURES
source
1..777
/organism="Plasmid P307"
/specific_host="Escherichia coli"
/db_xref="taxon:2472"
/clone="PAT153."
/plasmid="Plasmid P307"
1..777
/gene="lta"
1..777
/gene="lta"
1..777
/codon_start=1
/transl_table=11
/product="heat-labile enterotoxin subunit A"
/protein_id="AAB98202.1"
/db_xref="GI:150459"
/translation="MKNTFFIFILLASPLVANGDRLYRADSRPDEIKRSGGLMPRG

CDS
/strain="CH-C2"
/serotype="O-136"
/specific_host="pig"
/db_xref="taxon:562"
/country="Taiwan"
/notes="Isolated from hemolytic cells"
1..777
/contig="ltpa"
/codon_start=1
/transl_table=11
/product="heat-labile enterotoxin subunit A precursor"
/protein_id="AAL5671.1"
/db_xref="GI:18026886"
/translation="MKNTFFIFILLASPLVANGDRLYRADSRPDEIKRSGGLMPRG
HNEYFDRTGMNINLYDHARGTQTFVRVDDGYSTSLRSASHLAGOSILSGSYTY
IYVATAPMFMFNVNDLVGYSPHPYEQEVSALGGLPYQIYGMVYVNFVDERLARN
REYRDRIYRNINLAPAEDEGRLAGPPDQANREEPWIHHAQGGCNSSRTTGDTCN
EETQNLSTIYLEYOSKVRQIFSDYQSEVDIYNIRDEL"
1..54
sig_peptide
55..630
mat_peptide
/product="heat-labile enterotoxin subunit A1"
637..774
mat_peptide
/product="heat-labile enterotoxin subunit A2"
136 c 164 g 222 t

BASE COUNT 255 a 136 c 164 g 222 t
ORIGIN

Alignment Scores:
Pred. No.: 5.38e-07 Length: 777
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

SEQ1-65TO79 (1-15) x AF242417 (1-777)
QY 1 SerleuArgSerAlaHisLeuAlaGlyGlnSerIleuSergly 15
|||||
Db 247 AGTTTGAGAGTGTCTCATTAGCAGCAGCTATATTATCAGCA 291

RESULT 4
LOCUS A04913 777 bp DNA linear PAT 13-JUL-1993
DEFINITION E. coli LTA gene for labile toxin A.
ACCESSION A04913
VERSION A04913.1 GI:412215
KEYWORDS labile toxin A.
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 777)
AUTHORS Hayes,M.V., Harford,S. and Ross,G.W.
TITLE Microbiological Process
JOURNAL Patent: EP 0145486-A 5 19-JUN-1985;
GLAXO GROUP LIMITED
FEATURES
source
1..777
/organism="Escherichia coli"
1..777
/db_xref="taxon:562"
1..777
/codon_start=1
/transl_table=11
/product="labile toxin A"
/protein_id="CAA00402.1"
/db_xref="GI:412215"
/db_xref="SWISS-PROT:P06717"
/translation="MKNTFFIFILLASPLVANGDRLYRADSRPDEIKRSGGLMPRG
HNEYFDRTGMNINLYDHARGTQTFVRVDDGYSTSLRSASHLAGOSILSGSYTY
IYVATAPMFMFNVNDLVGYSPHPYEQEVSALGGLPYQIYGMVYVNFVDERLARN
REYRDRIYRNINLAPAEDEGRLAGPPDQANREEPWIHHAQGGCNSSRTTGDTCN
EETQNLSTIYLEYOSKVRQIFSDYQSEVDIYNIRDEL"
1..6
sig_peptide
6..630
mat_peptide
/product="heat-labile enterotoxin subunit A1"
637..774
mat_peptide
/product="heat-labile enterotoxin subunit A2"
136 c 164 g 222 t

BASE COUNT 255 a 136 c 164 g 222 t
ORIGIN

Alignment Scores:
Pred. No.: 5.38e-07 Length: 777
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

SEQ1-65TO79 (1-15) x A04913 (1-777)
QY 1 SerleuArgSerAlaHisLeuAlaGlyGlnSerIleuSergly 15
|||||
Db 247 AGTTTGAGAGTGTCTCATTAGCAGCAGCTATATTATCAGCA 291

RESULT 5
LOCUS A04915 777 bp DNA linear PAT 14-JUL-1993
```

```

HNEYFDRTGMNINLYDHARGTQTFVRVDDGYSTSLRSASHLAGOSILSGSYTY
IYVATAPMFMFNVNDLVGYSPHPYEQEVSALGGLPYQIYGMVYVNFVDERLARN
REYRDRIYRNINLAPAEDEGRLAGPPDQANREEPWIHHAQGGCNSSRTTGDTCN
EETQNLSTIYLEYOSKVRQIFSDYQSEVDIYNIRDEL"
1..54
sig_peptide
55..630
mat_peptide
/product="heat-labile enterotoxin subunit A1"
637..774
mat_peptide
/product="heat-labile enterotoxin subunit A2"
136 c 164 g 222 t

BASE COUNT 255 a 136 c 164 g 222 t
ORIGIN

Alignment Scores:
Pred. No.: 5.38e-07 Length: 777
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

SEQ1-65TO79 (1-15) x A04913 (1-777)
QY 1 SerleuArgSerAlaHisLeuAlaGlyGlnSerIleuSergly 15
|||||
Db 247 AGTTTGAGAGTGTCTCATTAGCAGCAGCTATATTATCAGCA 291

RESULT 5
LOCUS A04915 777 bp DNA linear PAT 14-JUL-1993
```

```

DEFINITION Artificial sequence LTA gene for labile toxin A.
ACCESSION A04915
VERSION A04915.1 GI:412519
KEYWORDS labile toxin A.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 777)
AUTHORS Hayes,M.V., Harford,S. and Ross,G.W.
TITLE Microbiological process
JOURNAL Patent: EP 0145486-A 7 19-JUN-1985;
GLAXO GROUP LIMITED
FEATURES
    source
        1..777
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        1..777
        /codon_start=1
        /transl_table=11
        /product="labile toxin A"
        /protein_id="CAA00403.1"
        /db_xref="GI:412520"
        /translation="MKNTFFIFILLASPLYANGDRLYRADSRPDEIKRSGGLMPRG
        HNEYPDRGTQMNINLYDHARCTQGFVATDGYVTSLSLSRAHLAAGOSTILSGYSTY
        IYVATAPNMFNNDVGLGVSPHPYEDVSALGGIPYSQIYGMVRNFGVIDERLARN
        REYRDYRYRNLIAPAEDGYRLAGPPDQAWREEPWIHAHPQGCGNSRRTITGDTCN
        EETONLSTIYLREYQSKVKRQIFSDYQSEVDIYNRIRDEL"
BASE COUNT 255 a 135 c 164 g 222 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 5.38e-07 Length: 777
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
SEQ1-65TO79 (1-15) x A04915 (1-777)
QY 1 SerLeuArGSeRaLaHIsLeuAlaGlyInSeRtIleuSeRgly 15
Db 247 AGTTGAGAGAGTGCCTACCTTAGCAGACAGCTATATTATTCAGGA 291
RESULT 6
LOCUS ENSLTPA 866 bp DNA linear BCT 26-APR-1993
DEFINITION Plasmid ENT p307 (enterotoxigenic E.coli) LTP subunit A gene.
ACCESSION M15362
VERSION M15362.1 GI:148337
KEYWORDS LTP gene; heat-labile toxin.
SOURCE Plasmid ENT (enterotoxigenic E.coli porcine isolate) DNA, clone p370.
ORGANISM Plasmid ENT
REFERENCE 1 (bases 1 to 866)
AUTHORS Yamamoto,T., Gojobori,T. and Yokota,T.
TITLE Evolutionary origin of pathogenic determinants in enterotoxigenic
JOURNAL J. Bacteriol. 169 (3), 1352-1357 (1987)
MEDLINE 87137303
PUBMED 3546273
COMMENT Clean copy of sequence [1] kindly provided by T.Yamamoto
(22-APR-1987).
FEATURES
    source
        1..866
        /organism="Plasmid ENT"
        /db_xref="taxon:2463"
        /plasmid="Plasmid ENT"
        90..866
        /note="Heat-labile toxin subunit A precursor"
        /codon_start=1
        /transl_table=11

```

```

        /protein_id="AAA24793.1"
        /db_xref="GI:148338"
        /translation="MKNTFFIFILLASPLYANGDRLYRADSRPDEIKRSGGLMPRG
        HNEYPDRGTQMNINLYDHARCTQGFVATDGYVTSLSLSRAHLAAGOSTILSGYSTY
        IYVATAPNMFNNDVGLGVSPHPYEDVSALGGIPYSQIYGMVRNFGVIDERLARN
        REYRDYRYRNLIAPAEDGYRLAGPPDQAWREEPWIHAHPQGCGNSRRTITGDTCN
        EETONLSTIYLREYQSKVKRQIFSDYQSEVDIYNRIRDEL"
        90..143
        /note="Heat-labile toxin subunit A signal peptide"
        mat_peptide
            144..863
            /product="Heat-labile toxin subunit A"
BASE COUNT 277 a 149 c 181 g 259 t
ORIGIN Unreported.
Alignment Scores:
Pred. No.: 5.91e-07 Length: 866
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
SEQ1-65TO79 (1-15) x ENSLTPA (1-866)
QY 1 SerLeuArGSeRaLaHIsLeuAlaGlyInSeRtIleuSeRgly 15
Db 336 AGTTGAGAAAGTGCCTACCTTAGCAGACAGCTATATTATTCAGGA 380
RESULT 7
LOCUS ENRLTPA 898 bp DNA linear BCT 26-APR-1993
DEFINITION Plasmid ENT-R PCG86 (enterotoxigenic E.coli) LTP subunit A gene.
ACCESSION M15361
VERSION M15361.1 GI:148333
KEYWORDS LTP gene; heat-labile toxin.
SOURCE Plasmid ENT-R (enterotoxigenic E.coli porcine isolate) DNA, clone PCG86.
ORGANISM Plasmid ENT-R
REFERENCE 1 (bases 1 to 898)
AUTHORS Yamamoto,T., Gojobori,T. and Yokota,T.
TITLE Evolutionary origin of pathogenic determinants in enterotoxigenic
JOURNAL J. Bacteriol. 169 (3), 1352-1357 (1987)
MEDLINE 87137303
PUBMED 3546273
COMMENT Clean copy of sequence [1] kindly provided by T.Yamamoto
(22-APR-1987).
FEATURES
    source
        1..898
        /organism="Plasmid ENT-R"
        /db_xref="taxon:2462"
        /plasmid="Plasmid ENT-R"
        122..898
        /note="Heat labile toxin subunit A precursor"
        /codon_start=1
        /transl_table=11
        /protein_id="AAA24791.1"
        /db_xref="GI:148334"
        /translation="MKNTFFIFILLASPLYANGDRLYRADSRPDEIKRSGGLMPRG
        HNEYPDRGTQMNINLYDHARCTQGFVATDGYVTSLSLSRAHLAAGOSTILSGYSTY
        IYVATAPNMFNNDVGLGVSPHPYEDVSALGGIPYSQIYGMVRNFGVIDERLARN
        REYRDYRYRNLIAPAEDGYRLAGPPDQAWREEPWIHAHPQGCGNSRRTITGDTCN
        EETONLSTIYLREYQSKVKRQIFSDYQSEVDIYNRIRDEL"
        122..175
        /note="Heat labile toxin subunit A signal peptide"
        mat_peptide
            176..895
            /product="Heat labile toxin subunit A"
BASE COUNT 289 a 155 c 186 g 268 t
ORIGIN Unreported.
Alignment Scores:
Pred. No.: 6.1e-07 Length: 898

```


Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

SEQ1-65TO79 (1-15) x ENRLTPA (1-898)

OY 1 SerleuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
|||||

DB 368 AGTTTGAGAAAGTGTCTACTTAGCAGACAGCTATATTATTCAGCA 412

RESULT 8
ECOTOXA 934 bp DNA 1linear BCT 26-APR-1993
LOCUS E.coli toxA gene encoding subunit A of heat-labile enterotoxin.
ACCESSION K01995
VERSION K01995.1 GI:148027
KEYWORDS enterotoxin; heat-labile enterotoxin; toxA gene.
SOURCE E.coli H10407 (serotype 078:H11) plasmid DNA, clone pUY27.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 934)
AUTHORS Yamamoto,T., Tamura,T. and Yokota,T.
TITLE Primary structure of heat-labile enterotoxin produced by
Escherichia coli pathogenic for humans
JOURNAL J Biol. Chem. 259 (8), 5037-5044 (1984)
MEDLINE 84185610
PUBMED 6325417

COMMENT This heat-labile enterotoxin is pathogenic for humans (Lth).
Compared in [1] with LTP A and CT A (Ltp-pathogenic for piglets,
CT-cholera toxin).

FEATURES
Source Location/Qualifiers
1..934
/organism="Escherichia coli"
/db_xref="taxon:562"
158..934
/note="heat-labile enterotoxin A prepeptide"
/codon_start=1
/transl_table=11
/protein_id="AA24685.1"
/db_xref="GI:148028"
/translation="MKNTFFIFILLAPLYANGDKLYRADSRPPEIKRSGGLMPRG
HNEYFDRGTOMNINLYDHARGTGTGFVDDGYVSTLSIRSAHLAGOSTLGSSTYY
IYIATAPKMNENNVDLYGAPSPHPQEVSAIGCIPIYSQIYGRVNRFGIDRLIRN
REYDRYIRRLNLAAPAEDEGTRLAGFPFDHQAEEPEWIIHAPVGGCDSSRTTGDTCN
EETQNLSTIYLRKYQSKVRKQIFSDYQSEVDIYNRIINEL."
158..211
/note="heat-labile enterotoxin A signal peptide"
212..787
mat_peptide /product="heat-labile enterotoxin A1"
794..931
mat_peptide /product="heat-labile enterotoxin A1"

BASE COUNT 298 a 165 c 186 g 285 t
ORIGIN 46 bp upstream of FokI recognition pattern.

Alignment Scores:
Pred. No.: 6.31e-07 Length: 934
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

SEQ1-65TO79 (1-15) x ECOTOXA (1-934)

OY 1 SerleuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
|||||

DB 404 AGTTTGAGAAAGTGTCTACTTAGCAGACAGCTATATTATTCAGCA 448

RESULT 9
E03421

LOCUS E03421 1148 bp DNA 1linear PAT 29-SEP-1997
DEFINITION DNA encoding Lth of entero toxigenic Escherichia coli.
ACCESSION E03421
VERSION E03421.1 GI:2171637
KEYWORDS JP 1992079898-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1148)
AUTHORS Danbara,H. and Abe,A.
TITLE DNA AND RNA PROBE TO SIMULTANEOUSLY DETECT VIBRIO CHOLERAE AND
ENTEROTOXIGENIC ESCHERICHIA COLI AND DETECTION OF VIBRIO CHOLERA
AND ENTEROTOXIGENIC ESCHERICHIA COLI USING THE SAME DNA AND RNA
PROBE
JOURNAL Patent: JP 1992079898-A 1 13-MAR-1992;
KITASATO INST:THE
COMMENT OS Escherichia coli
PM JP 1992079898-A/1
PD 13-MAR-1992
PF 23-JUL-1990 JP 1990194208
PI DANBARA HIROFUMI, ABE AKIO
PC C12Q1/68,C07H21/02,C07H21/04,C12N15/11,C12N15/31,C12Q1/04, PC
C12Q1/10,
PC (C12N15/31,C12R1:19),(C12Q1/04,C12R1:63,C12R1:19): CC
strandedness: Double:
CC topology: Linear:
CC hypothetical: No;
CC anti-sense: No;
CC *source: clone=PKAD008;
FH Key Location/Qualifiers
FH misc_feature 1..1148
FT /note="DNA fragment encoding Lth of entero
FT toxigenic
FT Escherichia coli'
FT sig_peptide 1..54
FT mat_peptide 55..774
FT /product="subunit A of Lth"
FT 1..777
FT CDS /product="precursor subunit A of Lth" FT
FT sig_peptide 774..837
FT mat_peptide 838..1145
FT /product="subunit B of Lth"
FT CDS 774..1148
FT /product="precursor of subunit B of Lth".
FEATURES
Source Location/Qualifiers
1..1148
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 399 a 205 c 229 g 315 t
ORIGIN

Alignment Scores:
Pred. No.: 7.56e-07 Length: 1148
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

SEQ1-65TO79 (1-15) x E03421 (1-1148)

OY 1 SerleuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
|||||

DB 247 AGTTTGAGAAAGTGTCTACTTAGCAGACAGCTATATTATTCAGCA 291

RESULT 10
S60731 1275 bp DNA 1linear BCT 29-JAN-2002
LOCUS S60731
DEFINITION heat-labile enterotoxin A subunit, heat-labile enterotoxin B
subunit [Escherichia coli, 21d, Genomic, 1275 nt].
ACCESSION S60731

VERSION S60731.1 GI:408994
KEYWORDS
SOURCE Escherichia coli 21d.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1275)
AUTHORS Inoue,T., Tsuji,T., Koto,M., Imamura,S. and Miyama,A.
TITLE Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic Escherichia coli is identical to that of human strain H 10407
JOURNAL FEMS Microbiol. Lett. 108 (2), 157-161 (1993)
MEDLINE 93252225
PUBMED 8486242
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI gibanq 131491] from the original journal article.
This sequence comes from Fig. 2.
FEATURES
source
1..1275
/organism="Escherichia coli"
/db_xref="taxon:562"
68..844
/note="This sequence comes from Fig. 2; conceptual translation presented here differs from translation in publication; Ltc A subunit"
/codon_start=1
/transl_table=1
/product="heat-labile enterotoxin A subunit"
/protein_id="AAC6040.2"
/db_xref="GI:18390349"
/translation="MKNTFFIFILLASPLYANGDKLYRADSRPPDEIKRSGGLMPRG
HNEFDRTGOMINILYDHARGTQTFVRYDDGYVSTLSLSAHLAGOSILSGSTYY
IVYIATPMPFNVDNLGYSPHPYEDVSAAGCIPYSGIYGMVRYNREVIDERLHRN
REYRDYRNLNLIAPEADGYRLAGFPDPOHAMEEPEWIIHAPQGCGSSRTITDTCN
EETQNLSTLYLKRKQSKVKRQIFSDYQSEVDIYNIRNDEL"
841..1215
/note="This sequence comes from Fig. 2; Ltc B subunit"
/codon_start=1
/transl_table=1
/product="heat-labile enterotoxin B subunit"
/protein_id="AAC6041.1"
/db_xref="GI:408996"
/translation="MNKVCYVLFTHALLSLCAYGAPQSITELCSEYRNTQIYTTNDK
ILSYTESMAGKREMYIITFKSGATFQVEVPGSQHDSOKAIERMKDYLRIITYLETETK
IDKLCVNNKTPNSIAISMEN"
BASE COUNT 436 a 227 c 248 g 364 t
ORIGIN
Alignment Scores:
Pred. No.: 8.28e-07 Length: 1275
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
SEQ1-65TO79 (1-15) x S60731 (1-1275)
QY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSerGly 15
Db 314 AGTTTGAGAGAGTGTCTACCTAGCAGACAGCTATATTATCAGGA 358
RESULT 11
LOCUS AB011677 1434 bp DNA linear BCT 23-APR-1998
DEFINITION Escherichia coli genes for heat-labile enterotoxin A subunit and B subunit, complete cds.
ACCESSION AB011677
VERSION AB011677.1 GI:3062900
KEYWORDS heat-labile enterotoxin B subunit; Lth b subunit; heat-labile enterotoxin A subunit; Lth a subunit.
SOURCE Escherichia coli (strain:1032 (Enterotoxigenic)) DNA.
ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Tamura,S., Asanuma,H., Tomita,T., Komase,K., Kawahara,K.,
Dandara,H., Hattori,N., Watanabe,K., Suzuki,T., Nagamine,T.,
Aizawa,C., Oya,A. and Kurata,T.
TITLE Escherichia coli heat-labile enterotoxin B subunits supplemented with a trace amount of the holotoxin as an adjuvant for nasal influenza vaccine
JOURNAL Vaccine 12 (12), 1083-1089 (1994)
MEDLINE 95091056
REFERENCE 2 (bases 1 to 1434)
AUTHORS Komase,K.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1998) Katsuhiko Komase, The Kitasato Institute,
Center for Basic Research, Lab. of Virology; Shirokane 5-9-1,
Minato-ku, Tokyo 108-8642, Japan (E-mail:komase-kek@kitasato.or.jp,
Tel:+81-3-3444-6161, Fax:+81-3-3444-6637)
FEATURES
source
1..1434
/organism="Escherichia coli"
/strain="1032 (Enterotoxigenic)"
/db_xref="taxon:562"
91..867
/gene="Lth a subunit"
91..867
/gene="Lth b subunit"
/codon_start=1
/transl_table=1
/product="heat-labile enterotoxin A subunit"
/protein_id="BA25725.1"
/db_xref="GI:3062901"
/translation="MKNTFFIFILLASPLYANGDKLYRADSRPPDEIKRSGGLMPRG
HNEFDRTGOMINILYDHARGTQTFVRYDDGYVSTLSLSAHLAGOSILSGSTYY
IVYIATPMPFNVDNLGYSPHPYEDVSAAGCIPYSGIYGMVRYNREVIDERLHRN
REYRDYRNLNLIAPEADGYRLAGFPDPOHAMEEPEWIIHAPQGCGSSRTITDTCN
EETQNLSTLYLKRKQSKVKRQIFSDYQSEVDIYNIRNDEL"
864..1238
/gene="Lth b subunit"
864..1238
/gene="Lth a subunit"
/codon_start=1
/transl_table=1
/product="heat-labile enterotoxin B subunit"
/protein_id="BA25726.1"
/db_xref="GI:3062902"
/translation="MNKVCYVLFTHALLSLCAYGAPQSITELCSEYRNTQIYTTNDK
ILSYTESMAGKREMYIITFKSGATFQVEVPGSQHDSOKAIERMKDYLRIITYLETETK
IDKLCVNNKTPNSIAISMEN"
BASE COUNT 488 a 258 c 286 g 402 t
ORIGIN
Alignment Scores:
Pred. No.: 9.17e-07 Length: 1434
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
SEQ1-65TO79 (1-15) x AB011677 (1-1434)
QY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSerGly 15
Db 337 AGTTTGAGAGAGTGTCTACCTAGCAGACAGCTATATTATCAGGA 381
RESULT 12
LOCUS A16419 711 bp DNA linear PAT 06-JUL-1995
DEFINITION A subunit of a labile toxin LTI.
ACCESSION A16419
VERSION A16419.1 GI:641008
KEYWORDS

```

SOURCE          Escherichia coli.
ORGANISM        Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
FEATURES
  source        1..711
                /organism="Escherichia coli"
                /db_xref="taxon:562"
  gene          1..711
                /gene="LT1 subunit A"
                <1..711
  CDS           /gene="LT1 subunit A"
                /codon_start=1
                /transl_table=11
                /protein_id="CA01380.1"
                /db_xref="GI:641009"
                /translation="NGDRLYRADSRPDEIKRFSRMPGNEYFDRGTOMINLYDHA
                RGTOTGFVRDGYVSTSLRSALHAGYILSGYSLTIYVIANFNVNDIVSYSP
                HPEGEVSALGIPYSQIYGWRVNFVGVIDERLHRREYRDYRNLNAPADGYRL
                AGPPPDHQAAREPPIHHAPOGCGSSRTITGDTCMETQNLSTIYLRYSQSVKRNQI
                FSDYQSEVDIYRIRDEL"
BASE COUNT     234 a 123 c 156 g 198 t
ORIGIN
Alignment Scores:
Pred. No.:      0.101      Length:      711
Score:          10.00      Matches:      10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    66.67%      Indels:      0
DB:             6          Gaps:          0

SEQ1-65TO79 (1-15) x A16419 (1-711)
QY 1 SerleuArgSerAlaHisLeuAlaGlyGln 10
|||||
Db 190 AGTTTGAGAGTGTCTCATTAGCAGCAGAC 219

RESULT 13
AR118597
LOCUS          AR118597 711 bp DNA linear PAT 16-MAY-2001
DEFINITION     Sequence 5 from patent US 6149919.
ACCESSION      AR118597
VERSION        AR118597.1 GI:14100507
KEYWORDS
SOURCE         Unknown.
ORGANISM       Unclassified.
REFERENCE      1 (bases 1 to 711)
AUTHORS       Domenighini, M., Rappuoli, R., Pizza, M. and Hol, W.
TITLE         Immunogenic detoxified mutants of cholera toxin and of the toxin
              LT, their preparation and their use for the preparation of vaccines
JOURNAL        Patent: US 6149919-A 5 21-NOV-2000;
FEATURES
  source        1..711
                /organism="unknown"
BASE COUNT     234 a 123 c 156 g 198 t
ORIGIN
Alignment Scores:
Pred. No.:      0.101      Length:      711
Score:          10.00      Matches:      10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    66.67%      Indels:      0
DB:             6          Gaps:          0

SEQ1-65TO79 (1-15) x AR118597 (1-711)
QY 1 SerleuArgSerAlaHisLeuAlaGlyGln 10
|||||
Db 190 AGTTTGAGAGTGTCTCATTAGCAGCAGAC 219

```

```

RESULT 14
ECELTA
LOCUS          ECELTA 868 bp DNA linear BCT 12-SEP-1993
DEFINITION     E. coli gene elcA encoding the A subunit of the heat-labile
                enterotoxin (LT).
ACCESSION      V00275 J01604
VERSION        V00275.1 GI:41339
KEYWORDS       enterotoxin; signal peptide.
SOURCE         Escherichia coli.
ORGANISM       Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE      1 (bases 1 to 868)
AUTHORS       Spicer, E.K. and Noble, J.A.
TITLE         Escherichia coli heat-labile enterotoxin. Nucleotide sequence of
              the A subunit gene
JOURNAL        J. Biol. Chem. 257 (10), 5716-5721 (1982)
MEDLINE        82167425
PUBMED         6279611
FEATURES
  source        1..868
                /organism="Escherichia coli"
                /db_xref="taxon:562"
                /product="messenger RNA"
  CDS           /codon_start=1
                /transl_table=11
                /product="enterotoxin subunit A"
                /protein_id="CA23532.1"
                /db_xref="GI:41340"
                /db_xref="SWISS-PROT:P06717"
                /translation="MKNTFFIFILLASPLANGDRLYRADSRPDEIKRFSRMPG
                NEYFDRGTOMINLVDHARGTQGFVRDGYVSTSLRSALHAGYILSGYSLTIY
                IVIANFNVNDIVSYSPHPEGEVSALGIPYSQIYGWRVNFVGVIDERLHRREYR
                DRYRNLNINAPADGYRLAGPPPDHQAAREPPIHHAPOGCGSSRTITGDTCMETQ
                NLSTIYLRYSQSVKRNQIFSDYQSEVDIYRIRDEL"
BASE COUNT     284 a 142 c 179 g 263 t
ORIGIN
Alignment Scores:
Pred. No.:      0.12      Length:      868
Score:          10.00      Matches:      10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    66.67%      Indels:      0
DB:             1          Gaps:          0

SEQ1-65TO79 (1-15) x ECELTA (1-868)
QY 1 SerleuArgSerAlaHisLeuAlaGlyGln 10
|||||
Db 333 AGTTTGAGAGTGTCTCATTAGCAGCAGAC 362

RESULT 15
E00613
LOCUS          E00613 777 bp DNA linear PAT 29-SEP-1997
DEFINITION     DNA encoding A component of heat labile toxin, LTa.
ACCESSION      E00613
VERSION        E00613.1 GI:2168892
KEYWORDS       JP 1986005097-A/1.
SOURCE         Escherichia coli.
ORGANISM       Escherichia coli.
REFERENCE      1 (bases 1 to 777)
AUTHORS       Maikuru, B.H., Suteilibun, H. and Goodon, U.R.
TITLE         MICROBIAL METHOD
JOURNAL        Patent: JP 1986005097-A 1 10-JAN-1986;
COMMENT        GILAXO GROUP LTD
                OS Escherichia coli
                PN JP 1986005097-A/1

```



```

* 6533 7728: contig of 1196 bp in length
* 7729 7828: gap of unknown length
* 7829 9076: contig of 1248 bp in length
* 9077 9176: gap of unknown length
* 9177 10389: contig of 1213 bp in length
* 10390 10489: gap of unknown length
* 10490 11716: contig of 1227 bp in length
* 11717 11816: gap of unknown length
* 11817 13068: contig of 1252 bp in length
* 13069 13168: gap of unknown length
* 13169 14248: contig of 1080 bp in length
* 14249 14348: gap of unknown length
* 14349 15422: contig of 1074 bp in length
* 15423 15532: gap of unknown length
* 15523 16544: contig of 1022 bp in length
* 16545 16644: gap of unknown length
* 16645 18723: contig of 2079 bp in length
* 18724 18824: gap of unknown length
* 18824 19997: contig of 1174 bp in length
* 19998 20097: gap of unknown length
* 20098 21201: contig of 1104 bp in length
* 21202 22399: gap of unknown length
* 22399 23039: contig of 1638 bp in length
* 23040 24948: gap of unknown length
* 24948 25048: contig of 1909 bp in length
* 25049 26822: contig of 1774 bp in length
* 26823 26922: gap of unknown length
* 26923 28710: contig of 1788 bp in length
* 28711 28810: gap of unknown length
* 28811 30546: contig of 1736 bp in length
* 30547 30646: gap of unknown length
* 30647 31921: contig of 1275 bp in length
* 31922 32021: gap of unknown length
* 32022 34036: contig of 2015 bp in length
* 34037 34136: gap of unknown length
* 34137 36081: contig of 1945 bp in length
* 36082 36181: gap of unknown length
* 36182 38565: contig of 2384 bp in length
* 38566 38665: gap of unknown length
* 38666 41309: contig of 2644 bp in length
* 41310 41409: gap of unknown length
* 41410 43181: contig of 1772 bp in length
* 43182 43281: gap of unknown length
* 43282 46680: contig of 3399 bp in length
* 46681 46780: gap of unknown length
* 46781 49043: contig of 2263 bp in length
* 49044 49143: gap of unknown length
* 49144 51108: contig of 1965 bp in length
* 51109 51208: gap of unknown length
* 51209 53460: contig of 2252 bp in length
* 53461 53560: gap of unknown length
* 53561 57650: contig of 4090 bp in length
* 57651 57750: gap of unknown length
* 57751 60055: contig of 2305 bp in length
* 60056 60155: gap of unknown length
* 60156 63979: contig of 3824 bp in length
* 63980 64079: gap of unknown length
* 64080 67093: contig of 3014 bp in length
* 67094 67193: gap of unknown length
* 67194 69652: contig of 2459 bp in length
* 69653 69752: gap of unknown length
* 69753 72007: contig of 2255 bp in length
* 72008 72107: gap of unknown length
* 72108 73676: contig of 1569 bp in length
* 73677 73776: gap of unknown length
* 73777 78342: contig of 4566 bp in length
* 78343 78442: gap of unknown length
* 78443 82905: contig of 4463 bp in length
* 82906 83005: gap of unknown length
* 83006 86533: contig of 3548 bp in length
* 86534 90794: gap of unknown length
* 90794: contig of 4141 bp in length

```

```

* 90795 90894: gap of unknown length
* 90895 94509: contig of 3615 bp in length
* 94510 94610: gap of unknown length
* 94610 100334: contig of 5725 bp in length
* 100335 100434: gap of unknown length
* 100435 100466: contig of 5032 bp in length
* 105467 105566: gap of unknown length
* 105567 111968: contig of 6402 bp in length
* 111969 112068: gap of unknown length
* 112069 117777: contig of 5709 bp in length
* 117778 117877: gap of unknown length
* 117878 125639: contig of 7762 bp in length
* 125640 131285: gap of unknown length
* 131286 131385: contig of 5546 bp in length
* 131386 138670 138769: gap of unknown length
* 138670 147046: contig of 8277 bp in length
* 147047 147146: gap of unknown length
* 147147 155416: contig of 8270 bp in length.

FEATURES
    source
        1..153416
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="CH230-398M12"

BASE COUNT  47324 a 28148 c 27100 g 46106 t 6738 others

Alignment Scores:
Pred. No.: 126 Length: 155416
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 60.00% Indels: 0
Db: 2 Gaps: 0

SEQ1-65TO79 (1-15) x AC114463 (1-155416)
QY 1 SerLeuArGSeRAlhSLeuAlaGly 9
Db 96759 AGTCTCGAGGAGTCCACCTTGCTGGG 96733

RESULT 17
AC120962/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-27017, *** SEQUENCING IN PROGRESS
ACCESSION
AC120962
VERSION
AC120962.2 GI:21902870
KEYWORDS
HTG: HTGS PHASE1.
SOURCE
Rattus norvegicus.
ORGANISM
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 174342)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alstrooms,S.L., Amaratunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbarta,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Checko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,J., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsel,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,

```

```

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounsbred,H.,
Lozado,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okunolu,C.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatok,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 174342)
Worley,K.C.
Direct Submission
Submitted (14-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 174342)
Worley,K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20564310.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWMGD
Center clone name: CH230-27017
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 11189 bases at least Q40
Consensus quality: 11825 bases at least Q30
Consensus quality: 12214 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafile.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 84 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1411: contig of 1411 bp in length
* 1412 1511: gap of unknown length
* 1512 3172: contig of 1661 bp in length
* 3173 3272: gap of unknown length
* 3273 4708: contig of 1436 bp in length
* 4709 4808: gap of unknown length
* 4809 5969: contig of 1161 bp in length
* 5970 6069: gap of unknown length
* 6070 7232: contig of 1163 bp in length
* 7233 7332: gap of unknown length
*
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounsbred,H.,
Lozado,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okunolu,C.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatok,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 174342)
Worley,K.C.
Direct Submission
Submitted (14-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 174342)
Worley,K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20564310.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWMGD
Center clone name: CH230-27017
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 11189 bases at least Q40
Consensus quality: 11825 bases at least Q30
Consensus quality: 12214 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafile.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 84 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1411: contig of 1411 bp in length
* 1412 1511: gap of unknown length
* 1512 3172: contig of 1661 bp in length
* 3173 3272: gap of unknown length
* 3273 4708: contig of 1436 bp in length
* 4709 4808: gap of unknown length
* 4809 5969: contig of 1161 bp in length
* 5970 6069: gap of unknown length
* 6070 7232: contig of 1163 bp in length
* 7233 7332: gap of unknown length
*
7333 8481: contig of 1149 bp in length
8482 8581: gap of unknown length
8582 9766: contig of 1185 bp in length
9767 9865: gap of unknown length
9866 11203: contig of 1337 bp in length
11204 11303: gap of unknown length
11304 12664: contig of 1361 bp in length
12665 12764: gap of unknown length
12765 13983: contig of 1219 bp in length
13984 14083: gap of unknown length
14084 15688: contig of 1606 bp in length
15689 15789: gap of unknown length
15790 16933: contig of 1142 bp in length
16932 17031: gap of unknown length
17032 18402: contig of 1371 bp in length
18403 18502: gap of unknown length
18503 19681: contig of 1179 bp in length
19682 19781: gap of unknown length
19782 21173: contig of 1392 bp in length
21174 21273: gap of unknown length
21274 22487: contig of 1214 bp in length
22488 22587: gap of unknown length
22588 24267: contig of 1680 bp in length
24268 24367: gap of unknown length
24368 25985: contig of 1618 bp in length
25986 26085: gap of unknown length
26086 27515: contig of 1430 bp in length
27516 27615: gap of unknown length
27616 29062: contig of 1447 bp in length
29063 29162: gap of unknown length
29163 30220: contig of 1058 bp in length
30221 30320: gap of unknown length
30321 31547: contig of 1227 bp in length
31548 31647: gap of unknown length
31648 33327: contig of 1680 bp in length
33328 33427: gap of unknown length
33428 35365: contig of 1938 bp in length
35366 35465: gap of unknown length
35466 37641: contig of 2176 bp in length
37642 37741: gap of unknown length
37742 39281: contig of 1340 bp in length
39282 39381: gap of unknown length
39382 40849: contig of 1468 bp in length
40850 40949: gap of unknown length
40950 42588: contig of 1639 bp in length
42589 42688: gap of unknown length
42689 43879: contig of 1191 bp in length
43880 43979: gap of unknown length
43980 45235: contig of 1256 bp in length
45236 45335: gap of unknown length
45336 47109: contig of 1774 bp in length
47110 47209: gap of unknown length
47210 49153: contig of 1944 bp in length
49154 49253: gap of unknown length
49254 50744: contig of 1490 bp in length
50745 52280: gap of unknown length
52281 52380: contig of 1437 bp in length
52381 53563: gap of unknown length
53564 53663: contig of 1183 bp in length
53664 55365: gap of unknown length
55366 55465: contig of 1702 bp in length
55466 57862: gap of unknown length
57863 57962: contig of 2397 bp in length
57963 59274: gap of unknown length
59275 59374: contig of 1312 bp in length
59375 60518: gap of unknown length
60519 61825: contig of 1144 bp in length
61826 61925: gap of unknown length
61926 63355: contig of 1207 bp in length
63356 63455: gap of unknown length
63456 64936: contig of 1430 bp in length
64937 64936: contig of 1481 bp in length

```

```

*      64937      65036: gap of unknown length
*      65037      66641: contig of 1605 bp in length
*      66741      66642: gap of unknown length
*      66742      68059: contig of 1318 bp in length
*      68060      68159: gap of unknown length
*      68160      70194: contig of 2035 bp in length
*      70294      70294: gap of unknown length
*      70295      71882: contig of 1588 bp in length
*      71883      71982: gap of unknown length
*      71983      73578: contig of 1596 bp in length
*      73579      73678: gap of unknown length
*      73679      75687: contig of 2009 bp in length
*      75688      75787: gap of unknown length
*      75788      77065: contig of 1278 bp in length
*      77066      77165: gap of unknown length
*      77166      79166: contig of 2001 bp in length
*      79167      79266: gap of unknown length
*      79267      81213: contig of 1947 bp in length
*      81214      81313: gap of unknown length
*      81314      83169: contig of 1856 bp in length
*      83170      83269: gap of unknown length
*      83270      84818: contig of 1549 bp in length
*      84819      84918: gap of unknown length
*      84919      86110: contig of 1192 bp in length

```

Alignment Scores:

```

Pred. No.:      139      length:      174342
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      60.00%      Indels:      0
DB:              2      Gaps:          0

```

SEQ1-65TO79 (1-15) x AC120962 (1-174342)

QY 1 SerLeuArgSerAlaHisLeuAlaGly 9

DB 3585 AGTCTCAGGAGTGCACCTGCGCGG 3559

RESULT 18

```

LOCUS      BC021349      1877 bp      mRNA      linear      ROD 07-AUG-2002
DEFINITION Mus musculus, clone MGC:29285 IMAGE:3985822, mRNA, complete cds.
ACCESSION      BC021349
VERSION      BC021349.1 GI:18204168
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1877)
Strausberg, R.
Direct Submission
Submitted (14-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@hghrl.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

```

REFERENCE

AUTHORS
TITLE
JOURNAL

REMARK

COMMENT

FEATURES

source

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
Series: IRAC Plate: 39 Row: b Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.

Location/Qualifiers

CDS

```

1..1877
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="MGC:29285 IMAGE:3985822"
/tissue_type="mammary tumor. WAF-trf alpha model. 7 months
old, gross tissue."
/clone_11b="NCI-CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
154..1653
/codon_start=1
/product="Unknown (protein for MGC:29285)"
/protein_id="AAH21349.1"
/db_xref="GI:18204169"
/translation="MKRRDLPPSDATYATFALFNVAESPWKDSALQSLKRLROQARN
FOINIKYTHALLKVAKCADLRICDVRKEITROGHAVTEPFCLVIGCIDDKRTG
ROMAVMROVQAKTVHGSLSRHYEALEROLFEPSSKLGSPALPAPARTSRQPEVT
TAEPCHTALPPLAPKPTLLEVSLSIGTSPVVSFGVATPADRLALMGLEGF
LAKMTEHGLAPDIPKTLTLAEVEPESAESSLSVLDNRHREADVTFENLIRKSK
LGDLEGAKALPDLIAKKGIYPRVLCNLAICGHRPROMOHLADMKSOYSPNHIY
SLTINAAIKKIDPITYLITIKMRONSVNVNRYVROLEFFAEVPTPDRIYKGNITL
EKIDFPRAYIKWMLKAMPAAEPHPHMQEPQKRPVSDQDTTKAGGLRDC"

```

BASE COUNT 473 a 504 c 525 g 375 t

ORIGIN

Alignment Scores:

```

Pred. No.:      31      length:      1877
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      53.33%      Indels:      0
DB:              10      Gaps:          0

```

SEQ1-65TO79 (1-15) x BC021349 (1-1877)

QY 1 SerLeuArgSerAlaHisLeuAla 8

DB 353 AGCCTGAGTCTGCACACTTGCGG 330

RESULT 19

```

LOCUS      BC004766      3014 bp      mRNA      linear      ROD 07-AUG-2002
DEFINITION Mus musculus, similar to KIA0632 protein, clone MGC:7042
IMAGE:3156304, mRNA, complete cds.
ACCESSION      BC004766
VERSION      BC004766.1 GI:13435830
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3014)
Strausberg, R.
Direct Submission
Submitted (21-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk

```

REFERENCE

AUTHORS
TITLE
JOURNAL

REMARK

COMMENT

Email: c9apbs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amgdbcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hultk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 6 Row: 0 Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES

source

```
1..3014
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:7042 IMAGE:3156304"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
684..2183
/codon_start=1
/product="Similar to K1MA0632 protein"
/protein_id="AA04766.1"
/db_xref="GI:13435831"
/translation="MKRRDLESDATYATLFWNCAESPWKDSALQSLKRLQOLQARN
FQNLKTYHALKYNACADRLCLIDYFKELIQGHAVTEETFCFLVGCIDOKTGF
RQAMQWOMSLGKIKPSRNGVNLLEAARDGSGDEVASRLITSEENILLPKPK
GRHMAGRVQAKTVGWSLRVEALEROLFLEPSOKLEGPALPEARVTSIOPEVT
TAEGHTVALPLAPKPTHTLESLISLISPAVYSGVATPADRLMLGLSEF
LGMTHHGLDPDITLITLLAVYVPGSAESSLISVLDNRHVEDVTFNTLLKRSK
LGLDEGAKALPLTAKKGIENLRTFCNALGCHRPDGMQLADMKSOVSPNTHY
STLINALAKTLDYLLISILKDMKONSVPVNEVVVROLFEAAEYPTFDRYKGNLYL
EKIDGFRVRYQWMLKAMPAAEAPHPMQEFOFNKPVGDODTTDKAGLRDG"
```

CDS

```
BASE COUNT      755 a      833 c      843 g      583 t
ORIGIN
Alignment Scores:
Pred. No.:      46.9      Length:      3014
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      53.33%      Indels:      0
DB:              10      Gaps:      0
```

SEQ1-65TO79 (1-15) x BC004766 (1-3014)

OY 1 SerleuArgSerAlaHisIleuAla 8
 |||||||

Db 883 ACCCTGAGGCTGCACACTTGCG 860

RESULT 20

HSU10336 4011 bp mRNA linear INV 08-OCT-1994
 LOCUS HSU10336/c
 DEFINITION Helobdella stagnalis RNA polymerase II largest subunit mRNA,
 partial cds.

ACCESSION U10336

VERSION U10336.1 GI:520512

KEYWORDS

Helobdella stagnalis.
 Helobdella stagnalis
 Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;
 Rhynchobdellida; Glossiphoniidae; Helobdella.

REFERENCE

1 (bases 1 to 4011)
 AUTHORS Sidow, A. and Thomas, W.K.
 TITLE A molecular evolutionary framework for eukaryotic model organisms
 JOURNAL Curr. Biol. 4 (7), 596-603 (1994)
 MEDLINE 95041334
 PUBMED 7953533
 REFERENCE 2 (bases 1 to 4011)
 AUTHORS Sidow, A.
 TITLE Direct Submission
 JOURNAL Submitted (03-JUN-1994) Arend Sidow, Molecular and Cell Biology,
 University of California, 401 Barker Hall, Berkeley, CA 94720, USA

FEATURES

source

```
1..4011
/organism="Helobdella stagnalis"
/db_xref="taxon:34588"
/dev_stage="early embryos"
/notes="Sequence obtained by PCR amplification with  

degenerate oligos (external to the sequence shown) from  

first strand cDNA. See Reference 1 for details"
<1..>4011
```

CDS

```
/codon_start=1
/product="RNA polymerase II, largest subunit"
/protein_id="AA50227.1"
/db_xref="GI:520513"
/translation="HIELSKPVHVPFFTKIIRLCVYCSKLLIDPNHPKVDIL
NKRNMGRGPGVNDICVKTSCGCGETDKDDNDPNIVGVCCGRHOPKIRNL
LEITPAKROINDNDQKRMILTAERLYLEFKATITBETCTILDMRKYARPMILITML
PVPPLPRPRAVWFGSAKNODDITRLALITIKINOLKRNENGAHILADITMLQ
FVAHAIIDNEMDPLPRAQKSGRPLKSIQKRGKRGRLGMKRGVDFSRVITP
DPMRLIDQGVPRSLAQNLTPEIYVPEIETELVKGAMQYGAKEILKETDRI
DLRYHPSSTDLHLCFGYRVERHMODNDVYIFNRQPLIKMSMKRVRKILPSTFRIL
LSVTPRYNADPDSDEMLLAOSLETRAISOLASVKKRMITPOANRPVMGIVDSLT
AVAKMTRPRTFTTKDEINIMLITLPRGMACKLQPMILKRALMTKOLFSTITPERIN
VIRHTSTHEDDEDGPHKWIPEDTLVVEDCKLSGLICKSLDASSGSDIHTHE
LGSDATADYAYIOWVTNMLLVGHTTIGIADTADAKYSOIQAIRKAKSDVEVI
EKANHDELEPMPEGNTLRQTFENOVNAILNDARDKGSLOKSLSEFNKSVVAGSK
GNKINISQVLAAGQOONVGRKILPEGRFORTLPHTIKDYGSESGFVENSYLALTP
TEEFFPAMGREGELDITAVKTAETGYIORRLIKAMSVKVKDGTVRNOIBDILRY
GEGCAGAEVVEQNDLPSTKPSNKAEPAGKPPDPTNEKLLKNTLDDDLKSLNGDANVI
AEVSEYKQLEDRATIROIPSGDSKYLPENLORLITMAOKITRITRKRNSHPV
KILIEDVRELSEKFMVKGEDRLKANTANATLNMNLVNSTCSRVLEELSTEAR
EWMGEIESKFTQARVQPEWGAALAOISGAPQOMLNTFHVAGSVKNTLVGPR
LKIIIVWSKPRAPSLTVLILIGOPADAKADAVLCQETLIRAVTENTATLYYDPD
MLHLEDEDEWYIYYDMPDVISRLSPMLLVEIDRKMTRKKTLMFOISKTIAGF
GDLQNCITENDNAEKLYLVRILMSNDGQODDTEQDIDKMPDQFTHIESNMILTM
TLOGITSIKAYNQOPTTDDKRIITIDEGEKRALQDMITLEDGTALRVLSEYVDP
VKTQVNDIVVEFEVDEGTEAVRKSIEREMNNVTSFGSVYVNRHLLALCDVMYAKGHM
ATRHGINRQETGVAR"
```

BASE COUNT 1256 a 820 c 920 g 1011 t 4 others

ORIGIN

Alignment Scores:
 Pred. No.: 60.1 Length: 4011
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 3 Gaps: 0

SEQ1-65TO79 (1-15) x HSU10336 (1-4011)

OY 7 leuAlaGlyInserIleleuSer 14
 |||||||

Db 2418 CTCGCCGGCAATCATCTCTCC 2395

RESULT 21

AE013887 9968 bp DNA linear BCT 26-JUL-2002
 LOCUS AE013887/c
 DEFINITION Yersinia pestis K1M section 287 of 415 of the complete genome.

ACCESSION AE013887

VERSION AE013887.1 GI:21959739

KEYWORDS

Yersinia pestis K1M.

ORGANISM	Yersinia pestis KIM	gene	complement(2817. .3947)
REFERENCE	Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae; Yersinia.	CDS	/gene="y2845"
AUTHORS	1 (bases 1 to 9968)		/complement(2817. .3947)
TITLE	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,		/gene="y2845"
JOURNAL	Liss, P., Pena, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C.,		/function="putative"
PUBMED	Fetherston, J.D., Lindler, L.E., Brubaker, R.R., Plana, G.V.,		/note="residues 1 to 372 of 376 are 74.73 pct identical to
AUTHORS	Stratley, S.C., McDonough, K.A., Nilles, M.L., Matson, J.S.,		residues 1 to 372 of 375 from E. coli K12 : B0859;
	Blattner, F.R. and Perry, R.D.		residues 1 to 372 of 376 are 75.53 pct identical to
	Genome Sequence of Yersinia pestis KIM		residues 1 to 372 of 375 from GenPept :
	J. Bacteriol. 184 (16), 4601-4611 (2002)		>gb AG55235.1 AE005266_14 (AE005266) putative enzyme
	12142430		[Escherichia coli O157:H7 EDL933]"
	2 (bases 1 to 9968)		/codon_start=1
	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,		/transl_table=11
	Liss, P., Pena, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C.,		/product="putative enzyme"
	Fetherston, J.D., Lindler, L.E., Brubaker, R.R., Plana, G.V.,		/protein_id="AA086396.1"
	Stratley, S.C., McDonough, K.A., Nilles, M.L., Matson, J.S.,		/db_xref="GI:21959742"
	Blattner, F.R. and Perry, R.D.		/translation="MHCAQITACRSCOMLDKPYRQOLADKOHLESLLAGHAYTQW
	Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445		LAVFGRSAFRNKAAMVSGSVREPLDMLHRDTPVDLCAQPLVPSPSEVETVLK
	Henry Mall, Madison, WI 53706, USA		TFIARAGLTPYVYARKRGELKLLLESTYNGELMPLVLRSETKLAOLIALPMLQO
FEATURES	Location/Qualifiers		OLPQLAVISANIQPVMAIIIEGREIPLTEOQALPRFNQVPLXIRPQSEFQNPQVA
source	1. .9968		ASLYATARQWQHEVHSMWDLFCGSGGHCAGCPETQLEIENAAEIAACROSAE
gene	/organism="Yersinia pestis KIM"		OLGKRVSAALDSTRFATAEADQIPELVYNPRRGIGRELDYLSOMAPKFLVSSC
	/db_xref="taxon:187410"		MetMKDLSLGAHIERVQLEDFMFPHTSHLEVLTLTR
	/gene="hflC"		complement(4031. .4495)
	/note="y2843"		/gene="y2846"
	/complement(214. .1140)		/complement(4031. .4495)
	/function="enzyme; macromolecule degradation: Degradation		/gene="y2846"
	of proteins, peptides, glyco"		/note="residues 6 to 142 of 154 are 42.44 pct identical to
	/note="residues 12 to 271 of 308 are 25.00 pct identical		residues 14 to 152 of 162 from E. coli K12 : B0858;
	to residues 3 to 247 of 283 from GenPept :		residues 1 to 142 of 154 are 42.36 pct identical to
	>gq AAD36886.1 AE001819_9 (AE001819) ftsH protease		residues 4 to 147 of 157 from GenPept : >emb CAD05320.1
	activity modulator HflC [thermotoga maritima]"		(Al627268) putative membrane protein [salmonella enterica
	/codon_start=1		subsp. enterica serovar Typh]"
	/transl_table=11		/codon_start=1
	/product="putative ftsH proteinase activity modulator"		/transl_table=11
	/protein_id="AA086394.1"		/product="hypothetical protein"
	/db_xref="GI:21959740"		/protein_id="AA086397.1"
	/translation="MTMKKEMSRFVYVCGFLAILTLIAVICLMSWYTNESDGI		/db_xref="GI:21959743"
	TKWGVAAVAEPLGFKIDITEETISINRSIKYDLKAYSKDQAPQMYVISGFQ		/translation="MRPSFSSAAFPVPVIGGTATIAIRIIGVLLVLAELGSGVSD
	VPTSVEDLFVKGSIONMAERLVSRRHVPQVNVGOYAVSAGVNRBDFTARVEE		FISDSQRMDTGILFASVYLLLLDLMKCVANHRNNRMYTLICOLIIIMYLLMAS
	LRVLKDEPLINVNINDELTGEGYEAAGEEEMKAEVNEKRMKLEETKTADIAI		LNWLEPDIRIEDSGAEIHSLLOKIPDVYITGLLFFPMHFPQSKMSV"
	EQARGSESOLSTAKIGAEKIKLMGAERENITLMGAAEAETIKLRADLKQNPILVE		complement(4620. .5465)
	LITAEKWNGLPQTMLPNSSVPFINAKNDGK"		/gene="y2847"
	1932. .2690		/gene="y2847"
gene	/gene="arlJ"		/complement(4620. .5465)
	/note="y2844"		/gene="poti"
	1932. .2690		/function="transport; transport of small molecules; amino
	/gene="arlJ"		acids, amines"
	/function="transport; transport of small molecules; amino		/note="residues 1 to 280 of 281 are 84.28 pct identical to
	acids, amines"		residues 1 to 280 of 281 from E. coli K12 : B0857;
	/note="residues 10 to 252 of 252 are 78.60 pct identical		residues 1 to 280 of 281 are 85.00 pct identical to
	to residues 1 to 243 of 243 from E. coli K12 : B0860;		residues 1 to 280 of 281 from GenPept : >gb AL119816.1
	residues 4 to 252 of 252 are 77.51 pct identical to		(AE008737) ABC superfamily (membrane), putrescine
	residues 12 to 260 of 260 from GenPept :		transporter [Salmonella typhimurium LT2]"
	>gq AG55239.1 AE005267_4 (AE005267) arginine 3rd		/codon_start=1
	transport system periplasmic binding protein [Escherichia		/transl_table=11
	coli O157:H7 EDL933]"		/product="putrescine ABC transport protein; permease"
	/codon_start=1		/protein_id="AA086398.1"
	/transl_table=11		/db_xref="GI:21959744"
	/product="arginine 3rd transport system periplasmic		/translation="MNNLEPVRSMRVRILITVGYEFLYAPMLMLVITYSNSKLVTW
	binding protein"		AGSTRTWYQLEFNDSSAMISAVGLSTIAASATMAVAVLITAAVWVRGFRGSGTG
	/protein_id="AA086395.1"		AFMLTAPLVMPDVTGSLILFLVAMGHAGAEKGMETIMLAHYTCGAYTVVIS
	/db_xref="GI:21959741"		SRMLRELDRISEFAAMDLAGTAPLKVFVYVTPMTPAPLISGWLAFSLDDLYIASFV
	/translation="MKNSDRSLMKKLLATLISGMAFSATAETLRFASATYPPPE		ACGAGATTPLVTVSSVRMGVNPETINLATLILVGVIGLFAMWIMSRKQURBLR
	SMDANNEIVGFMDLAKLCROMEANCTFNQAFDLSIPALKRKYDAVSGMDIRE		KAARS"
	RSKOVAFTQPYRANSAIVIAPKGRFSFADLKRRKIGMENGTHOKLDRKHEIQTV		complement(5462. .6427)
	SISYONAIIDLNGRIDVGTAVVAVNEMLKTNPNLASVGEHVTDPOYFGTGLGAV		/gene="poch"
	RPDNIALLTKLNKAIDAIVKADGYQAINDKWPQ"		/complement(5462. .6427)
			/gene="y2848"
			/note="y2848"
			/complement(5462. .6427)
			/gene="poch"
			/function="transport; transport of small molecules; amino
			acids, amines"
			/note="residues 18 to 318 of 321 are 82.05 pct identical

		to residues 14 to 314 of 317 from E. coli K12 : B0856"
gene	/codon_start=1 /transl_table=11 /product="putrescine permease of ABC transporter" /protein_id="AA086399.1" /db_xref="GI:21959745" /translation="MIPESTGAASEPIRSGVPVALIQRORMAHGRLVITAMYLIMILFLMPLPLEYFKTSLAEMVAWAVPPYDILTWTWDGKIDISLNLGNLYLLDDPLYTDRAVLQIAVANASTLCCLIGIPLANALAHSKPTNLTILLIFSWTSFTLRIVAKMKOILKNNGLIINNEFIWMTGIIIDPGLVIHTNLAYIGVVSYLPFNVLFYTALTRLDISVEALLDGARPFKEFTFSVIIPLTKGGIVAGSMLVFIPAWGEFVIPDLGGPSIMIGFILMOEFNNRDMPAPASAVAATVMVLLLAIPLMFHKHQKKDIGAR" complement(6452..7585) /gene="potC" /note="y2849" complement(6452..7585)	
CDS	/function="transport; transport of small molecules; amino acids, amines" /note="residues 1 to 375 of 377 are 86.40 pct identical to residues 28 to 402 of 404 from E. coli K12 : B0855" /codon_start=1 /transl_table=11 /product="ATP-binding component of putrescine ABC transport system" /protein_id="AA086400.1" /db_xref="GI:21959746" /translation="MNVDYIRPRPKRSOKVFPTLEIRNIITSFGQAADVDTNTIYKGETIALGASGCCKSTLLRMLAGFEOPGOVIDGDLSHPVRYPRIINNRPQSVALFPMTWEONIAIFGLKODKLPSNELKSRYAEMLTLTHMGDEFARKPHOSGORVARLARSAIKRRPKILLDEPMGALDKLRDRQLVELILEEYGATCYMVITHDOEAANTMAGRITALNRKEVQIEPELETIEHNSRFESAEIFGSVNFFEGVKERLDALVTDSPGLRHPLVSDVASDVDPVPLVALPERKMVCDDPVPDGCMFAAGEVHAIFYADLDSLHYHHKIHSQMLSAQIQMHGRYRKGMPTWCDEVOLCMDADSCVVGSS" 7591..7710 /gene="y2850" 7591..7710 /gene="y2850" /function="unknown" /codon_start=1 /transl_table=11 /product="hypothetical" /protein_id="AA086401.1" /db_xref="GI:21959747" /translation="MLSVKAEQQAAFNVRRNKREODEKLPAPTRINFITYNE" complement(7756..8982) /gene="poTF" /note="y2851"	
gene		
CDS		
Pred. No.: Score: Percent Similarity: Best Local Similarity:	133 8.00 100.00%	Length: 9968 Matches: 8 Conservative: 0 Mismatched: 0 Indels: 0 Gaps: 0
Query Match: DB:	53.33% 1	
SEQ1-65t079 (1-15) x AE013887 (1-9968)		
GZ	1 SerLeuArgSerAlaHisLeuAla 8 	
Dd	2856 TCACCTACGAGAAGTCACCCTGTCT 2833	
RESULT 22 AC004462		
LOCUS	AC004462	36021 bp DNA linear PRI 27-MAY-20000
DEFINITION	Homo sapiens Chromosome 22q11.2 Cosmid Clone 18c3 In DGCR Region,	
COMPLETE SEQUENCE.		
ACCESSION	AC004462 U30597	
VERSION	AC004462.3 GI:7122648	
KEYWORDS	HTG.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL	Eukaryota: Metazoa: Chordata: Craniata; Vertebrata: Euteleostomi, Mammalia: Eutheria, Primates: Catarrhini, Hominoidea, Homo. 1 (bases 1 to 36021) Roe, B. A. HTGS Submission Unpublished 2 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (20-MAR-1998) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 3 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (05-MAY-1998) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 4 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (20-JUN-1998) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 5 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (20-JUN-1998) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 6 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (30-OCT-1998) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 7 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (16-NOV-1998) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 8 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (26-NOV-1998) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 9 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (27-MAY-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 10 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (10-JUN-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 11 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (01-MAR-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 12 (bases 1 to 36021) Roe, B. A. Direct Submission Submitted (02-MAR-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA 13 (bases 1 to 36021)
--	--

AUTHORS Roe,B.A.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA

COMMENT
 On Mar 1, 2000 this sequence version replaced gi:4895277.
 Because these overlapping clones came from different libraries
 there are numerous overlapping instances of insertions, deletions, and single
 nucleotide polymorphisms in the overlapping regions below.
 AC004461(119f4) 38146 44873 (0) overlaps AC004462(18c3) 1 6728
 (29293) AC004462(18c3) 35015 36021 (0) overlaps AC004471(111f11) 1
 1006 (43290).

FEATURES
 source
 1. .36021
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22q11.2"
 /clone="18c3"

BASE COUNT 10719 a 8334 c 8023 g 8945 t

ORIGIN

Alignment Scores:
 Pred. No.: 406 length: 36021
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 9 Gaps: 0

SEQ1-65TO79 (1-15) x AC004462 (1-36021)

QY 1 SerLeuArgSerAlaHisLeuAla 8
 |||||
 Db 30258 TCTTTGAGAGTGGCCATCTGGCC 30281

RESULT 23
 AC130257
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-41414, *** SEQUENCING IN PROGRESS
 *** 25 unordered pieces.
 AC130257 51644 bp DNA linear HTG 09-AUG-2002
 Rattus norvegicus
 HTG: HTGS-PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 51644)

REFERENCE
 AUTHORS
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alshbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbieri,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Burch,P., Burkelt,C., Burrell,K.L., Byrd,N.C.,
 Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaq,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hoggins,M., Holloway,C., Hollins,B.,
 Homs,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Lonsheed,H.,
 Lozadó,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 Submitted (09-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GYOZ
 Center clone name: CH230-41414
 ----- Summary Statistics
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 26108 bases at least Q40
 Consensus quality: 29630 bases at least Q30
 Consensus quality: 31890 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1607 1606: contig of 1606 bp in length
 1707 1706: gap of unknown length
 1707 3110: contig of 1404 bp in length
 3111 3210: gap of unknown length
 3211 4341: contig of 1131 bp in length
 4342 4441: gap of unknown length
 4442 5548: contig of 1107 bp in length
 5549 5649: gap of unknown length
 5649 6905: contig of 1257 bp in length
 6906 7006: gap of unknown length
 7006 8372: contig of 1367 bp in length
 8373 8472: gap of unknown length
 8473 10277: contig of 1805 bp in length
 10278 10377: gap of unknown length
 10378 12045: contig of 1666 bp in length
 12046 12145: gap of unknown length
 12146 13770: contig of 1625 bp in length
 13771 13870: gap of unknown length
 13871 16665: contig of 2795 bp in length
 16666 16766: gap of unknown length
 16766 18217: contig of 1452 bp in length
 18218 18318: gap of unknown length
 18318 20766: contig of 2449 bp in length

```

* 20767 20866: gap of unknown length
* 20867 22035: contig of 1169 bp in length
* 22036 22135: gap of unknown length
* 22136 23915: contig of 1780 bp in length
* 23916 24015: gap of unknown length
* 24016 25271: contig of 1256 bp in length
* 25272 25371: gap of unknown length
* 25372 28545: contig of 3174 bp in length
* 28546 28645: gap of unknown length
* 28646 30117: contig of 1472 bp in length
* 30118 30217: gap of unknown length
* 30218 32310: contig of 2093 bp in length
* 32311 32410: gap of unknown length
* 32411 33826: contig of 1416 bp in length
* 33827 33926: gap of unknown length
* 33927 35868: contig of 1942 bp in length
* 35869 35968: gap of unknown length
* 35969 39176: contig of 3208 bp in length
* 39177 39276: gap of unknown length
* 39277 41239: contig of 1963 bp in length
* 41240 41339: gap of unknown length
* 41340 45088: contig of 3749 bp in length
* 45089 45188: gap of unknown length
* 45189 48347: contig of 3159 bp in length
* 48348 48447: gap of unknown length
* 48448 51644: contig of 3197 bp in length.

FEATURES
  source      1..51644
              /organism="Rattus norvegicus"
              /db_xref="taxon:10116"
              /clone="CH230-41414"

BASE COUNT   13445 a 11551 c 11353 g 12853 t 2442 others
ORIGIN
Alignment Scores:
  Pred. No.:      556      Length:      51644
  Score:          8.00      Matches:      8
  Percent Similarity: 100.00%      Conservative: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match:      53.33%      Indels:      0
  DB:                2      Gaps:          0

Seq01-65to79 (1-15) x AC130257 (1-51644)
Qy      2      LeuArgSerAlaHisLeuAlaGly 9
Db      48619 CTTAGATCAGCTCATCTCCAGGC 48642

RESULT 24
AC111766
LOCUS      AC111766
DEFINITION Rattus norvegicus clone CH230-12122, *** SEQUENCING IN PROGRESS
ACCESSION AC111766.2 GI:21736856
VERSION   HTG; HTGS_PHASEL.
KEYWORDS  HTG; HTGS_PHASEL.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
           1 (bases 1 to 72519)
REFERENCE 1 (bases 1 to 72519)
AUTHORS  Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
          Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
          Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
          Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryan,N.P.,
          Buhay,C., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
          Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
          Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
          Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
          Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

```

```

          Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
          Falls,T., Fariaguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
          Homsl,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kratovic,J., Kureshi,A., Landry,N., Deal,B., Lewis,L.C., Lewis,L.,
          Li,J., Li,Z., Lichatage,O., Lien,C., Liu,J., Liu,W., Louisedge,H.,
          Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
          Meshawari,M., Mapa,P., Martin,R., Martindale,A., Martinez,B.,
          Messey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mel,G., Metzger,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,N., Nguyen,N.,
          Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
          Oraguine,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
          Scherer,S., Scott,G., Shen,H., Shooshari,N., Stinson,I.,
          Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
          Sulton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
          Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
          Umanai,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
          Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
          Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G., and Gibbs,R.
          Direct Submission
          Unpublished
          2 (bases 1 to 72519)
REFERENCE 1 (bases 1 to 72519)
AUTHORS  Worley,K.C.
TITLE    Direct Submission
JOURNAL  Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          3 (bases 1 to 72519)
REFERENCE 1 (bases 1 to 72519)
AUTHORS  Worley,K.C.
TITLE    Direct Submission
JOURNAL  Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Jul 12, 2002 this sequence version replaced gi:18701608.

COMMENT
  Center: Baylor College of Medicine
  Center code: BCM
  Web site: http://www.hgsc.bcm.tmc.edu/
  Contact: hgsc-help@bcm.tmc.edu
  Project Information
  Center project name: GOB0
  Center clone name: CH230-12122
  Summary Statistics
  Sequencing vector: Plasmid;
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.990329
  Consensus quality: 38289 bases at least Q40
  Consensus quality: 41688 bases at least Q30
  Consensus quality: 44356 bases at least Q20
  -----
  * NOTE: Estimated insert size may differ from sequence length
  * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 37 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 1003: contig of 1003 bp in length
  * 1004 1103: gap of unknown length
  * 1104 2183: contig of 1080 bp in length

```

```

* 2184 2283: gap of unknown length
* 2284 3795: contig of 1512 bp in length
* 3796 3895: gap of unknown length
* 3896 5066: contig of 1171 bp in length
* 5067 5167: gap of unknown length
* 5167 6361: contig of 1195 bp in length
* 6362 6461: gap of unknown length
* 6462 7735: contig of 1274 bp in length
* 7736 7835: gap of unknown length
* 7836 9226: contig of 1391 bp in length
* 9227 9326: gap of unknown length
* 9327 10778: contig of 1452 bp in length
* 10779 10878: gap of unknown length
* 10879 11987: contig of 1109 bp in length
* 11988 12087: gap of unknown length
* 12088 13862: contig of 1775 bp in length
* 13863 13962: gap of unknown length
* 13963 15237: contig of 1275 bp in length
* 15238 15337: gap of unknown length
* 15338 16402: contig of 1065 bp in length
* 16403 16502: gap of unknown length
* 16503 18904: contig of 2402 bp in length
* 18905 19004: gap of unknown length
* 19005 21067: contig of 2063 bp in length
* 21068 21167: gap of unknown length
* 21168 22261: contig of 1094 bp in length
* 22262 22361: gap of unknown length
* 22362 23828: contig of 1467 bp in length
* 23829 23928: gap of unknown length
* 23929 25798: contig of 1870 bp in length
* 25799 25898: gap of unknown length
* 25899 27126: contig of 1228 bp in length
* 27127 27227: gap of unknown length
* 27228 29396: contig of 2170 bp in length
* 29397 29496: gap of unknown length
* 29497 30944: contig of 1448 bp in length
* 30945 31044: gap of unknown length
* 31045 32283: contig of 1239 bp in length
* 32284 32383: gap of unknown length
* 32384 33814: contig of 1431 bp in length
* 33815 33914: gap of unknown length
* 33915 34933: contig of 1019 bp in length
* 34934 35033: gap of unknown length
* 35034 36594: contig of 1561 bp in length
* 36595 37885: gap of unknown length
* 37886 37985: gap of unknown length
* 37986 40387: gap of unknown length
* 40388 40487: gap of unknown length
* 40489 42974: contig of 2487 bp in length
* 42975 43074: gap of unknown length
* 43075 45379: contig of 2305 bp in length
* 45380 45479: gap of unknown length
* 45480 47495: contig of 2016 bp in length
* 47496 47595: gap of unknown length
* 47596 49855: contig of 2260 bp in length
* 49856 49955: gap of unknown length
* 49956 53360: contig of 3405 bp in length
* 53361 53460: gap of unknown length
* 53461 56411: contig of 2951 bp in length
* 56412 56511: gap of unknown length
* 56512 60105: contig of 3594 bp in length
* 60106 60205: gap of unknown length
* 60206 63910: contig of 3705 bp in length
* 63911 64011: gap of unknown length
* 64012 66282: contig of 2282 bp in length
* 66283 66392: gap of unknown length
* 66393 68650: contig of 2258 bp in length
* 68651 72519: gap of unknown length
* 72520 72519: contig of 3769 bp in length.

```

FEATURES

source

/organism="Rattus norvegicus"

```

/db_xref="taxon:10116"
/clone="CH230-12122"
BASE COUNT 19098 a 15658 c 15483 g 18574 t 3706 others
ORIGIN
Alignment Scores:
Pred. No.: 747
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 53.33%
DB: 2
Gaps: 0
SEQ1-65to79 (1-15) x AC111766 (1-72519)
OY 1 SerLeuArgSerAlaHisLeuAla 8
Db 67869 AGCCTCAGAGTCCACCTGACC 67892
RESULT 25
AC115213
LOCUS
DEFINITION
AC115213 81938 bp DNA linear HMG 18-JUL-2002
Rattus norvegicus clone CH230-105P11, *** SEQUENCING IN PROGRESS
***, 13 unordered pieces.
ACCESSION
AC115213
VERSION
AC115213.2 GI:21745935
KEYWORDS
HMG: HMG5.PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 81938)
Muzny,D.M., Adams,C., Adio-oduola,B., All-ouman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barberia,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowles,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,F., Garza,N., Gill,R.,
Gorell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsbom,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Manua,P., Martin,R., Martindale,A., Martinez,F.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolle,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vlnson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Wortley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 81938)
AUTHORS Morley, K.C.
JOURNAL Direct Submission
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 81938)
Morley, K.C.
REFERENCE Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:19482300.

COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTRJ
Center clone name: CH230-105P11
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 73691 bases at least Q40
Consensus quality: 74751 bases at least Q30
Consensus quality: 75325 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1233: contig of 1233 bp in length
* 1234 1333: gap of unknown length
* 1334 2438: contig of 1105 bp in length
* 2439 2538: gap of unknown length
* 2539 5723: contig of 3185 bp in length
* 5724 5823: gap of unknown length
* 5824 9892: contig of 4069 bp in length
* 9893 9993: gap of unknown length
* 9993 15300: contig of 5308 bp in length
* 15301 15400: gap of unknown length
* 15401 20646: contig of 5246 bp in length
* 20647 20746: gap of unknown length
* 20747 26462: contig of 5716 bp in length
* 26463 26562: gap of unknown length
* 26563 32260: contig of 5698 bp in length
* 32261 32360: gap of unknown length
* 32361 41018: contig of 8658 bp in length
* 41019 41118: gap of unknown length
* 41119 48816: contig of 7698 bp in length
* 48817 48916: gap of unknown length
* 48917 59153: contig of 10237 bp in length
* 59154 59253: gap of unknown length
* 59254 72159: contig of 12906 bp in length
* 72160 72259: gap of unknown length
* 72260 81938: contig of 9679 bp in length.
Location/Qualifiers
1. 81938
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-105P11"

BASE COUNT 23795 a 15659 c 15274 g 24926 t 2284 others
ORIGIN

Alignment Scores:
Pred. No.: 831 length: 81938
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

SEQ1-65to79 (1-15) x AC115213 (1-81938)

OY 1 SerleuArgSerAlaHisLeuAla 8
Db 72723 AGCCTCAGAGTGCACCTGACC 72746

RESULT 26
AC120099
LOCUS
DEFINITION Rattus norvegicus clone CH230-328D17, *** SEQUENCING IN PROGRESS
AC120099 83292 bp DNA linear HTG 23-JUL-2002
VERSION
KEYWORDS HTG; HTGS; PHASE1.
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 83292)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaralunga, H.C., Aye, J.R., Ayele, M., Banks, T.,
Barbata, J.J., Beaton, J., Bimaga, K., Blackeburg, K., Bonin, D.,
Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Devilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dihn, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Einhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
Falls, T., Ferraguto, D., Flaag, N., Ford, J.J., Foster, P., Frantz, P.,
Gablis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hoque, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louleghed, H.,
Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, S.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwum, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojucokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
Schere, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tsaney, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 83292)
AUTHORS Morley, K.C.

TITLE Direct Submission
JOURNAL Submitted (03-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 83292)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 18, 2002 this sequence version replaced gi:20428712.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GWJ
Center clone name: CH230-328D17
Summary Statistics
Sequencing vector: piasmld:
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 55608 bases at least Q40
Consensus quality: 61281 bases at least Q30
Consensus quality: 64903 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1071: contig of 1071 bp in length
1072 1171: gap of unknown length
1172 2345: contig of 1174 bp in length
2345 2445: gap of unknown length
2445 3465: contig of 1020 bp in length
3465 3565: gap of unknown length
3565 4861: contig of 1296 bp in length
4861 4862: gap of unknown length
4862 6391: contig of 1430 bp in length
6391 6491: gap of unknown length
6491 7516: contig of 1025 bp in length
7516 7617: gap of unknown length
7617 8682: contig of 1066 bp in length
8682 8783: gap of unknown length
8783 10785: contig of 2003 bp in length
10785 10885: gap of unknown length
10885 11903: contig of 1018 bp in length
11903 12004: gap of unknown length
12004 13244: contig of 1241 bp in length
13244 13344: gap of unknown length
13344 14365: contig of 1021 bp in length
14365 14465: gap of unknown length
14465 15995: contig of 1530 bp in length
15995 16096: gap of unknown length
16096 17567: contig of 1472 bp in length
17567 17667: gap of unknown length
17667 19031: contig of 1364 bp in length
19031 19131: gap of unknown length
19131 20446: contig of 1315 bp in length
20446 20546: gap of unknown length
20546 22777: contig of 2231 bp in length
22777 22878: gap of unknown length
22878 24114: contig of 1237 bp in length
24114 24214: gap of unknown length
24214 26322: contig of 2108 bp in length
26322 26422: gap of unknown length

26423 28278: contig of 1856 bp in length
28278 28378: gap of unknown length
28378 30593: contig of 2215 bp in length
30593 30694: gap of unknown length
30694 32710: contig of 2017 bp in length
32710 32810: gap of unknown length
32810 34846: contig of 2036 bp in length
34846 34946: gap of unknown length
34946 37504: contig of 2558 bp in length
37504 37604: gap of unknown length
37604 37605: contig of 2865 bp in length
37605 40469: gap of unknown length
40469 40569: gap of unknown length
40569 43343: contig of 2774 bp in length
43343 43443: gap of unknown length
43443 47409: contig of 3966 bp in length
47409 47510: gap of unknown length
47510 50487: contig of 2978 bp in length
50487 50587: gap of unknown length
50587 54255: contig of 3668 bp in length
54255 54355: gap of unknown length
54355 58901: contig of 4546 bp in length
58901 59001: gap of unknown length
59001 61856: contig of 2855 bp in length
61856 61956: gap of unknown length
61956 65408: contig of 3452 bp in length
65408 65508: gap of unknown length
65508 74378: contig of 8870 bp in length
74378 74379: gap of unknown length
74379 74478: gap of unknown length
74478 83292: contig of 8814 bp in length.

FEATURES
source 1..83292
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-328D17"

BASE COUNT 22175 a 16753 c 16700 g 24388 t 3276 others
ORIGIN

Alignment Scores:
Score: No.: 843 83292
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 53.33%
DB: 2
Gaps: 0

SEQ1-65TO79 (1-15) x AC120099 (1-83292)
QY 1 SerleuArgSerAlaHisLeuAla 8
DB 49425 AGCCTCAGAGTGCCGCCACCTGGCC 49448

RESULT 27
AC121484/c 85022 bp DNA 1linear HTG 24-JUL-2002
LOCUS Rattus norvegicus clone CH230-155L16, *** SEQUENCING IN PROGRESS
DEFINITION
***, 45 unordered pieces.
ACCESSION AC121484 GI:21909076
VERSION AC121484.2
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 85022)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunga,H.C., Are,J.R., Ayala,M., Banks,T., Barbarella,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaey,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,E.,
 Homi,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.B.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Kaitson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
 Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoohatari,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Welstock,G. and Glibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 85022)
 Worley,K.C.
 Direct Submission
 Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 85022)
 Worley,K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 19, 2002 this sequence version replaced gi:20976436.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GYTT
 Center clone name: CH230-155L16
 ----- Summary Statistics
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 44173 bases at least Q40
 Consensus quality: 47255 bases at least Q30
 Consensus quality: 49138 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 45 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.


```

* 62147 62246: gap of unknown length
* 62247 64433: contig of 2187 bp in length
* 64434 64533: gap of unknown length
* 64534 66039: contig of 1506 bp in length
* 66040 66139: gap of unknown length
* 66140 68657: contig of 2518 bp in length
* 68658 68757: gap of unknown length
* 71429 71528: gap of 2671 bp in length
* 71529 74724: contig of 3196 bp in length
* 74725 74824: gap of unknown length
* 74825 78455: contig of 3531 bp in length
* 78456 81466: gap of unknown length
* 81467 81566: contig of 3011 bp in length
* 81567 85022: gap of unknown length
* 85022 85022: contig of 3456 bp in length.

FEATURES
Source
1. .85022
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-155L16"

BASE COUNT 22027 a 20072 c 18892 g 19379 t 4652 others

ORIGIN
Alignment Scores:
Pred. No.: 858 Length: 85022
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

SEQ1-65TO79 (1-15) x AC121484 (1-85022)
QY 1 SerleuargSerAlaHisLeuAla 8
Db 83568 AGCCTCAGAGTCCACCTGGCC 83545

RESULT 28
AC105557/c
LOCUS Rattus norvegicus clone CH230-223K22, *** SEQUENCING IN PROGRESS
DEFINITION Rattus norvegicus clone CH230-223K22, *** SEQUENCING IN PROGRESS
ACCESSION AC105557
VERSION AC105557.2 GI:21736422
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 93766)
Muzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooms,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barberia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delanaye,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Hollaway,C., Hollins,B.,
Homs,J., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louised,H.,
Lozadaro,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Mashedari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Masssey,E., Maxhiney,E., McLeod,M.P., Meador,M., Mel,G., Metzner,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunoye,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,
Peters,L., Pichens,R., Primus,E., Pu,L.L., Quiles,M., Ruiz,Y.,
Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 93766)
Worley,K.C.
Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 93766)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18092779.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNCM
Center clone name: CH230-223K22
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 55649 bases at least Q40
Consensus quality: 59489 bases at least Q30
Consensus quality: 62770 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1121: contig of 1121 bp in length
* 1122 1221: gap of unknown length
* 1222 1222: contig of 1086 bp in length
* 1223 2308: gap of unknown length
* 2308 2407: gap of unknown length
* 2408 3413: contig of 1006 bp in length
* 3414 3513: gap of unknown length
* 3514 4814: contig of 1301 bp in length
* 4815 4915: gap of unknown length
* 4915 6096: contig of 1182 bp in length
* 6097 6196: gap of unknown length
* 6197 7291: contig of 1095 bp in length
* 7292 7392: gap of unknown length
* 7392 8427: contig of 1036 bp in length

```

8428	8527:	gap of unknown length
8528	9757:	contig of 1230 bp in length
9758	9857:	gap of unknown length
9858	11261:	contig of 1404 bp in length
11262	11361:	gap of unknown length
11362	12995:	contig of 1634 bp in length
12996	13095:	gap of unknown length
13096	14320:	contig of 1225 bp in length
14321	14420:	gap of unknown length
14421	15516:	contig of 1196 bp in length
15517	15716:	gap of unknown length
15717	17376:	contig of 1660 bp in length
17377	17476:	gap of unknown length
17477	18653:	contig of 1177 bp in length
18654	18753:	gap of unknown length
18754	20110:	contig of 1357 bp in length
20111	20210:	gap of unknown length
20211	22273:	contig of 2063 bp in length
22274	22373:	gap of unknown length
22374	23556:	contig of 1283 bp in length
23557	23756:	gap of unknown length
23757	26068:	contig of 2312 bp in length
26069	26168:	gap of unknown length
26169	27830:	contig of 1662 bp in length
27831	27930:	gap of unknown length
27931	29140:	contig of 1210 bp in length
29141	29240:	gap of unknown length
29241	30435:	contig of 1195 bp in length
30436	30535:	gap of unknown length
30536	31677:	contig of 1142 bp in length
31678	31777:	gap of unknown length
31778	33446:	contig of 1569 bp in length
33447	33446:	gap of unknown length
33447	36030:	contig of 2584 bp in length
36031	36130:	gap of unknown length
36131	37532:	contig of 1402 bp in length
37533	37632:	gap of unknown length
37633	38952:	contig of 1330 bp in length
38953	39052:	gap of unknown length
39053	41114:	contig of 2362 bp in length
41115	41514:	gap of unknown length
41515	42558:	contig of 1044 bp in length
42559	42658:	gap of unknown length
42659	44487:	contig of 1829 bp in length
44488	44587:	gap of unknown length
44588	46058:	contig of 1471 bp in length
46059	46158:	gap of unknown length
46159	47912:	contig of 1754 bp in length
47913	48012:	gap of unknown length
48013	49336:	contig of 1334 bp in length
49337	49436:	gap of unknown length
49437	50773:	contig of 1337 bp in length
50774	50873:	gap of unknown length
50874	53618:	contig of 2745 bp in length
53619	53718:	gap of unknown length
53719	56466:	contig of 2748 bp in length
56467	56566:	gap of unknown length
56567	58229:	contig of 1663 bp in length
58230	58329:	gap of unknown length
58330	60637:	contig of 2308 bp in length
60638	62729:	gap of unknown length
62730	62829:	gap of unknown length
62829	65181:	contig of 2352 bp in length
65182	65281:	gap of unknown length
65282	67339:	contig of 2058 bp in length
67340	67439:	gap of unknown length
67440	70168:	contig of 2729 bp in length
70169	70268:	gap of unknown length
70269	72080:	contig of 1812 bp in length
72081	72180:	gap of unknown length
72181	74551:	contig of 2311 bp in length
74552	74651:	gap of unknown length

FEATURES	source	Location/Qualifiers
BASE COUNT	26308 a 17090 c 17144 g 25349 t 7875 others	
ORIGIN	/organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="CH230-223K22"	
Alignment Scores:		
Pred. No.:	934	Length: 93766
Score:	8.00	Matches: 8
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	53.33%	Indels: 0
DB:	2	Gaps: 0
SEQ1-657079 (1-15) x AC105557 (1-93766)		
Oy	1 SerLeuArgSerAlaHisLeuAla 8	
Db	86557 TCTTTGAGGCTGTGCTCACCTCGCT 86534	
RESULT 29		
AC11610/c		
LOCUS	AC11610	99537 bp DNA linear HTG 13-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-223N5. *** SEQUENCING IN PROGRESS	
ACCESSION	AC11610	
VERSION	AC11610.2 GI:21736174	
KEYWORDS	HTG: HTGS-PHASE1.	
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
AUTHORS	1 (bases 1 to 99537) Muney,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayelo,M., Banks,T., Babatia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Eantheart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotelo,T., Falls,T., Ferrando,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratovic,C., Kunesl,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loubege,H., Lozada,W.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mania,P., Martin,R., Martindale,A., Martinez,F.	

Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newtonson, J., Newtonson, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okwona, G., Oragunye, N., Oyiedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Petersen, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shooshbari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99537)
AUTHORS Worley, K.C.
JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 99537)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18701377.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GMSJ
Center clone name: CH230-233N5
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 44414 bases at least Q40
Consensus quality: 48280 bases at least Q30
Consensus quality: 51099 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1009: contig of 1009 bp in length
* 1010 1109: gap of unknown length
* 1110 2150: contig of 1041 bp in length
* 2250: gap of unknown length
* 2351 3434: contig of 1184 bp in length
* 3435 3534: gap of unknown length
* 3535 4841: contig of 1307 bp in length
* 4842 4941: gap of unknown length
* 4942 5974: contig of 1033 bp in length
* 5975 6074: gap of unknown length
* 6075 7612: contig of 1538 bp in length
* 7613 7712: gap of unknown length
* 7713 8809: contig of 1097 bp in length
* 8810 8909: gap of unknown length
* 8910 10640: contig of 1731 bp in length
* 10641 10740: gap of unknown length

10741 12267: contig of 1527 bp in length
* 12268 12367: gap of unknown length
* 12368 13688: contig of 1321 bp in length
* 13689 13788: gap of unknown length
* 13789 14916: contig of 1128 bp in length
* 14917 15016: gap of unknown length
* 15017 16382: contig of 1366 bp in length
* 16383 16482: gap of unknown length
* 16483 17661: contig of 1179 bp in length
* 17662 17761: gap of unknown length
* 17762 19054: contig of 1293 bp in length
* 19055 19154: gap of unknown length
* 19155 20437: contig of 1282 bp in length
* 20437 20536: gap of unknown length
* 20537 21675: contig of 1139 bp in length
* 21676 21775: gap of unknown length
* 21776 23029: contig of 1254 bp in length
* 23030 23129: gap of unknown length
* 23130 24670: contig of 1541 bp in length
* 24671 24770: gap of unknown length
* 24771 26366: contig of 1596 bp in length
* 26367 26466: gap of unknown length
* 26467 28255: contig of 1789 bp in length
* 28256 28355: gap of unknown length
* 28356 29559: contig of 1204 bp in length
* 29560 31424: gap of unknown length
* 31425 31524: gap of unknown length
* 31525 32755: contig of 1131 bp in length
* 32756 33987: contig of 1232 bp in length
* 33988 34087: gap of unknown length
* 34088 35876: contig of 1789 bp in length
* 35877 35976: gap of unknown length
* 35977 37666: contig of 1690 bp in length
* 37667 37766: gap of unknown length
* 37767 39416: contig of 1650 bp in length
* 39417 39516: gap of unknown length
* 39517 41754: contig of 2238 bp in length
* 41755 41854: gap of unknown length
* 41855 43110: contig of 1256 bp in length
* 43111 44717: gap of unknown length
* 44718 44817: contig of 1507 bp in length
* 44817 47129: gap of unknown length
* 47129 47229: contig of 2312 bp in length
* 47130 47229: gap of unknown length
* 47230 49202: contig of 1973 bp in length
* 49203 49302: gap of unknown length
* 49303 51253: contig of 1951 bp in length
* 51254 51353: gap of unknown length
* 51354 53325: contig of 1972 bp in length
* 53326 53425: gap of unknown length
* 53426 55241: contig of 1816 bp in length
* 55242 59028: gap of unknown length
* 59028 59128: contig of 3687 bp in length
* 59029 59128: gap of unknown length
* 59129 61269: contig of 2141 bp in length
* 61270 61369: gap of unknown length
* 61370 64073: contig of 2704 bp in length
* 64074 64173: gap of unknown length
* 64174 65676: contig of 1503 bp in length
* 65677 65776: gap of unknown length
* 65777 68907: contig of 3131 bp in length
* 68908 69007: gap of unknown length
* 69008 71405: contig of 2398 bp in length
* 71406 71505: gap of unknown length
* 71506 74192: contig of 2687 bp in length
* 74193 74292: gap of unknown length
* 74293 77173: contig of 2881 bp in length
* 77174 77273: gap of unknown length
* 77274 80506: contig of 3233 bp in length
* 80507 80606: gap of unknown length
* 80607 83811: contig of 3205 bp in length

```

FEATURES
  source
    1..99537
      /organism="Rattus norvegicus"
      /db_xref="taxon:10116"
      /clone="CH230-233N5"
BASE COUNT   28179 a 18020 c 16896 g 30031 t      6411 others
ORIGIN
Alignment Scores:
Pred. No.:      984      Length:      99537
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      53.33%      Indels:      0
DB:              2      Gaps:      0
SEQ1-65TO79 (1-15) x AC111610 (1-99537)
Qy      1 SerleuargSerAlahisLeuAla 8
Db 49823 AGCCTCAGAGTGCCTCCACCTGCC 49800
RESULT 30
AC113697/c AC113697 100786 bp DNA linear HTG 17-JUL-2002
LOCUS      Rattus norvegicus clone CH230-98H10, *** SEQUENCING IN PROGRESS
DEFINITION *** 50 unordered pieces.
ACCESSION  AC113697
VERSION     AC113697.4 GI:21744433
KEYWORDS   HTG; HTGS; PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 100786)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsdorfs,S.T., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barakat,J., Benton,J., Blmage,K., Blankenburg,K., Bonini,D.,
            Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douhwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
            Homi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,
            Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokwenko,S., Ogub,M., Okunolu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Plickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
            Rivas,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,

```

```

TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 100786)
Worley,K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 100786)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:20303291.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSAC
Center clone name: CH230-98H10
----- Summary Statistics
Sequencing vector: piasmtd.
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 52457 bases at least Q40
Consensus quality: 55266 bases at least Q30
Consensus quality: 57254 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1597: contig of 1597 bp in length
* 1598 1697: gap of unknown length
* 1598 2757: contig of 1060 bp in length
* 2758 2857: gap of unknown length
* 2858 4369: contig of 1512 bp in length
* 4370 4469: gap of unknown length
* 4470 6153: contig of 1684 bp in length
* 6154 6253: gap of unknown length
* 6254 7408: contig of 1155 bp in length
* 7409 7508: gap of unknown length
* 7509 8643: contig of 1135 bp in length
* 8644 8743: gap of unknown length
* 8744 10450: contig of 1707 bp in length
* 10451 10550: gap of unknown length
* 10551 12188: contig of 1639 bp in length
* 12190 12289: gap of unknown length
* 12290 13375: contig of 1086 bp in length
* 13376 13475: gap of unknown length
* 13476 14475: contig of 1000 bp in length
* 14476 14575: gap of unknown length
* 14576 15611: contig of 1036 bp in length
* 15612 15711: gap of unknown length
* 15712 15823: contig of 1112 bp in length

```

```

* 16824 16923: gap of unknown length
* 16924 18263: contig of 1340 bp in length
* 18264 18363: gap of unknown length
* 18364 19566: contig of 1203 bp in length
* 19567 19666: gap of unknown length
* 19667 20729: contig of 1063 bp in length
* 20730 20829: gap of unknown length
* 20830 21997: contig of 1168 bp in length
* 21998 22097: gap of unknown length
* 22098 23635: contig of 1538 bp in length
* 23636 23735: gap of unknown length
* 23736 25034: contig of 1299 bp in length
* 25035 25134: gap of unknown length
* 25135 26618: contig of 1484 bp in length
* 26619 26718: gap of unknown length
* 26719 27777: contig of 1059 bp in length
* 27778 27877: gap of unknown length
* 27878 29026: contig of 1149 bp in length
* 29027 29126: gap of unknown length
* 29127 30485: contig of 1359 bp in length
* 30486 30585: gap of unknown length
* 30586 32122: contig of 1537 bp in length
* 32123 32222: gap of unknown length
* 32223 33861: contig of 1639 bp in length
* 33862 33961: gap of unknown length
* 33962 35570: contig of 1609 bp in length
* 35571 35670: gap of unknown length
* 35671 37905: contig of 2235 bp in length
* 37906 38005: gap of unknown length
* 38006 39145: contig of 1140 bp in length
* 39146 39245: gap of unknown length
* 39246 41343: contig of 2098 bp in length
* 41344 41443: gap of unknown length
* 41444 43600: contig of 2157 bp in length
* 43601 43700: gap of unknown length
* 43701 45302: contig of 1602 bp in length
* 45303 45402: gap of unknown length
* 45403 46631: contig of 1229 bp in length
* 46632 46731: gap of unknown length
* 46732 48388: contig of 1657 bp in length
* 48389 48488: gap of unknown length
* 48489 49889: contig of 1401 bp in length
* 49890 49989: gap of unknown length
* 49990 51444: contig of 1455 bp in length
* 51445 51544: gap of unknown length
* 51545 53624: contig of 2080 bp in length
* 53625 53724: gap of unknown length
* 53725 56264: contig of 2540 bp in length
* 56265 56364: gap of unknown length
* 56365 57940: contig of 1576 bp in length
* 57941 58040: gap of unknown length
* 58041 59274: contig of 1234 bp in length
* 59275 59374: gap of unknown length
* 59375 61855: contig of 2481 bp in length
* 61856 61955: gap of unknown length
* 61956 65555: contig of 3600 bp in length
* 65556 66555: gap of unknown length
* 66556 68670: contig of 2915 bp in length
* 68671 68671: gap of unknown length
* 70755 70855: gap of unknown length
* 70856 73538: contig of 2683 bp in length
* 73539 73638: gap of unknown length
* 73639 76503: contig of 2865 bp in length
* 76504 76603: gap of unknown length
* 76604 78577: contig of 1974 bp in length
* 78578 78677: gap of unknown length
* 81388 81388: contig of 2711 bp in length
* 81488 81488: gap of unknown length
* 83980 83980: contig of 2492 bp in length
* 84080 84080: gap of unknown length
* 84081 88640: contig of 4560 bp in length
* 88641 88740: gap of unknown length

```

```

* 88741 93676: contig of 4936 bp in length
* 93677 93776: gap of unknown length
* 93777 100786: contig of 7010 bp in length.

FEATURES
    source
        1..100786
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="CH230-98H10"

BASE COUNT    26810 a 19798 c 19905 g 29236 t 5037 others
ORIGIN

Alignment Scores:
    Pred. No.:      995      Length:      100786
    Score:          8.00      Matches:      8
    Percent Similarity: 100.00%      Conservative: 0
    Best Local Similarity: 100.00%      Mismatches: 0
    Query Match:      53.33%      Indels:      0
    DB:                2      Gaps:      0

SEQ1-65TO79 (1-15) x AC113697 (1-100786)

QY      1 SerLeuArgSerAlaHisLeuAla 8
        |||||
Db 52577 AGTCACGAGTCCACCGGCC 52554

RESULT 31
AC107589
LOCUS
DEFINITION
    Rattus norvegicus clone CH230-5701, *** SEQUENCING IN PROGRESS ***
AC107589
ACCESSION
    AC107589.3 GI:21736999
VERSION
    HTG: HTGS.PHASE1.
KEYWORDS
    Norway rat.
SOURCE
    Rattus norvegicus
    Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
    Rattus.
REFERENCE
    1 (bases 1 to 105756)
    Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
    Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
    Barbieri,J., Benton,J., Bimaye,K., Blankenburg,K., Bonin,D.,
    Bouche,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
    Buhey,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
    Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
    Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
    Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
    Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
    Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
    Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
    Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
    Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
    Gabisi,A., Gao,J., Garcia,A., Garner,T., Hale,S., Hamilton,K.,
    Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
    Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
    Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
    Homs,F., Howard,S., Huber,J., Hulyk,S., Hune,J., Jackson,L.E.,
    Jacobson,B., Jia,T., Johnson,R., Jolivet,S., Joudah,S.,
    Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
    Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
    Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Louisedg,H.,
    Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
    Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,F.,
    Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
    Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
    Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
    Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,
    Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
    Peters,L., Pichens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
    Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
    Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
    Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
    Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

```

Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vanson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 105756)
Worley, K.C.
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 105756)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18846366.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNUM
Center clone name: CH230-5701
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 54794 bases at least Q40
Consensus quality: 58130 bases at least Q30
Consensus quality: 61124 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1526: contig of 1526 bp in length
* 1527 1626: gap of unknown length
* 1627 2877: contig of 1251 bp in length
* 2878 2977: gap of unknown length
* 2978 4046: contig of 1069 bp in length
* 4047 4146: gap of unknown length
* 4147 5332: contig of 1186 bp in length
* 5333 5432: gap of unknown length
* 5433 6731: contig of 1299 bp in length
* 6732 6831: gap of unknown length
* 6832 7877: contig of 1046 bp in length
* 7878 7977: gap of unknown length
* 7978 9501: contig of 1524 bp in length
* 9502 9601: gap of unknown length
* 9602 10770: contig of 1169 bp in length
* 10771 12345: gap of unknown length
* 12346 12445: contig of 1475 bp in length
* 12446 13766: gap of unknown length
* 13767 13866: contig of 1321 bp in length
* 13867 15408: gap of unknown length
* 15409 15508: contig of 1542 bp in length
* 15509 16929: gap of unknown length
* 16930 17029: contig of 1421 bp in length
* 17030 18146: gap of unknown length
* 18147 18246: gap of unknown length

* 18247 19808: contig of 1562 bp in length
* 19809 19908: gap of unknown length
* 19909 20999: contig of 1091 bp in length
* 21000 21099: gap of unknown length
* 21100 22991: contig of 1892 bp in length
* 22992 23091: gap of unknown length
* 23092 24599: contig of 1508 bp in length
* 24600 24699: gap of unknown length
* 24700 26570: contig of 1871 bp in length
* 26571 26670: gap of unknown length
* 26671 27949: contig of 1279 bp in length
* 27950 28049: gap of unknown length
* 28050 29651: contig of 1602 bp in length
* 29652 29751: gap of unknown length
* 29752 31417: contig of 1666 bp in length
* 31418 31517: gap of unknown length
* 31518 32791: contig of 1274 bp in length
* 32792 32891: gap of unknown length
* 32892 33960: contig of 1069 bp in length
* 33961 34060: gap of unknown length
* 34061 36253: contig of 2193 bp in length
* 36254 36353: gap of unknown length
* 36354 37808: contig of 1435 bp in length
* 37809 37908: gap of unknown length
* 37909 39394: contig of 1486 bp in length
* 39395 39494: gap of unknown length
* 39495 41082: contig of 1588 bp in length
* 41083 41183: gap of unknown length
* 41184 42846: contig of 1664 bp in length
* 42847 42946: gap of unknown length
* 42947 44419: contig of 1473 bp in length
* 44420 44519: gap of unknown length
* 44520 46079: contig of 1560 bp in length
* 46080 46180: gap of unknown length
* 46181 47897: contig of 1718 bp in length
* 47898 47997: gap of unknown length
* 47998 49514: contig of 1517 bp in length
* 49515 49614: gap of unknown length
* 49615 52486: contig of 2872 bp in length
* 52487 52586: gap of unknown length
* 52587 54483: gap of unknown length
* 54484 54583: contig of 1897 bp in length
* 54584 55765: contig of 1182 bp in length
* 55766 55865: gap of unknown length
* 55866 57818: contig of 1953 bp in length
* 57819 57918: gap of unknown length
* 57919 59307: contig of 1388 bp in length
* 59308 59407: gap of unknown length
* 59408 61592: contig of 2185 bp in length
* 61593 61692: gap of unknown length
* 61693 63090: contig of 1398 bp in length
* 63091 63190: gap of unknown length
* 63191 65670: contig of 2480 bp in length
* 65671 65770: gap of unknown length
* 65771 68360: contig of 2590 bp in length
* 68361 71226: gap of unknown length
* 71227 71326: contig of 2766 bp in length
* 71327 73222: gap of unknown length
* 73223 73322: contig of 1896 bp in length
* 73323 75737: gap of unknown length
* 75738 75837: contig of 2415 bp in length
* 75838 78019: gap of unknown length
* 78020 78120: contig of 2182 bp in length
* 78121 80525: gap of unknown length
* 80526 80625: contig of 2406 bp in length
* 80626 83148: gap of unknown length
* 83149 83248: contig of 2523 bp in length
* 83249 86911: gap of unknown length
* 86912 87010: contig of 3663 bp in length
* 87011 91666: gap of unknown length
* 91667 91766: contig of 4636 bp in length
* 91767 94787: contig of 3021 bp in length

```

FEATURES
  source
    1..105756
    Location/Qualifiers

Alignment Scores:
  Pired. No.:      1.04e+03      Length:      105756
  Score:           8.00          Matches:      8
  Percent Similarity: 100.00%    Conservative: 0
  Best Local Similarity: 100.00%  Mismatches:  0
  Query Match:      53.33%       Indels:       0
  DB:                2           Gaps:          0

SEQ1-65TO79 (1-15) x AC107589 (1-105756)
QY      1 SerleuargserAlahsleuAla 8
Db      64859 AGCCTCAGAGTGCCTGCGCC 64882

RESULT 32
AC094679      107637 bp      DNA      linear      HTG 20-DEC-2001
LOCUS      Rattus norvegicus clone CH230-5610, *** SEQUENCING IN PROGRESS ***
DEFINITION      61 unordered pieces.
AC094679      AC094679.2 GI:17941456
VERSION      HTG: HTGS_PHASE1.
KEYWORDS      Rattus norvegicus.
SOURCE      Rattus norvegicus.
ORGANISM      Rattus.
REFERENCE
  AUTHORS      1 (bases 1 to 107637)
    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
    Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarta,J.,
    Benton,T., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
    Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
    Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
    Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
    Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
    Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
    Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
    Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
    Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
    Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
    Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
    Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
    Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
    Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
    Hollins,B., Homs,F., Howard,S., Huber,J., Huliyk,S., Hume,J.,
    Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
    Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
    Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
    Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
    Lousegod,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
    Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
    Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
    Mel,O., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
    Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
    Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nkokenko,S.,
    Ogun,H., Okunolu,G., Oregunye,N., Oyiebo,R., Pace,A., Payton,B.,
    Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
    Quliz,M., Ren,Y., Rivers,M., Rojas,A., Rojubenkan,I., Rolfe,M.,
    Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,R., Shoshitari,N.,
    Slison,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
    Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
    Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellrodt,B., Thomas,N.,
    Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R.,
    Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

```

```

  TITLE
  JOURNAL
  REFERENCE
  AUTHORS
  TITLE
  JOURNAL

COMMENT
  Watlington,S., Williams,G., Williamson,A., Wlaczek,R., Wooden,S.,
  Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
  Weinstein,G. and Gibbs,R.
  Direct Submission
  Unpublished
  2 (bases 1 to 107637)
  Morley,K.C.
  Direct Submission
  Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On Dec 20, 2001 this sequence version replaced gi:15624514.
  ----- Genome Center
  Center: Baylor College of Medicine
  Center code: BCM
  Web site: http://www.hgsc.bcm.tmc.edu/
  Contact: hgsc-help@bcm.tmc.edu
  ----- Project Information
  Center project name: GBDY
  Center clone name: CH230-5610
  ----- Summary Statistics
  findPhrapList
  Consensus quality: 86463 bases at least Q40
  Consensus quality: 93299 bases at least Q30
  Consensus quality: 98313 bases at least Q20
  Estimated insert size: 70561; sum-of-contigs estimation
  Quality coverage: 0x in Q20 bases; agarose-tp estimation
  Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation
  -----
  * NOTE: Estimated insert size may differ from sequence length
  * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_delta.html).
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 61 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  1
  2132 2231: contig of 2131 bp in length
  2132 2231: gap of unknown length
  2232 6868: contig of 4637 bp in length
  6868 6868: gap of unknown length
  6869 8921: contig of 1953 bp in length
  8922 9021: gap of unknown length
  9022 11536: contig of 2515 bp in length
  11537 11636: gap of unknown length
  11637 13993: contig of 2357 bp in length
  13994 14093: gap of unknown length
  14094 17636: contig of 3543 bp in length
  17637 17736: gap of unknown length
  17737 20624: contig of 2888 bp in length
  20625 20724: gap of unknown length
  20725 22913: contig of 2189 bp in length
  22914 23013: gap of unknown length
  23014 24540: contig of 1527 bp in length
  24541 24640: gap of unknown length
  24641 26377: contig of 1737 bp in length
  26378 26477: gap of unknown length
  26478 28914: contig of 2437 bp in length
  28915 30861: gap of unknown length
  30862 30961: contig of 1847 bp in length
  30962 32983: gap of unknown length
  32983 33083: contig of 2022 bp in length
  33084 33393: gap of unknown length
  33393 35493: contig of 2310 bp in length
  35494 37776: gap of unknown length
  37776 37876: contig of 2283 bp in length
  37877 39223: gap of unknown length
  39223 39323: contig of 1347 bp in length
  39324 40816: contig of 1493 bp in length

```

```

* 40817 40916: gap of unknown length
* 40917 42984: contig of 2068 bp in length
* 42985 43084: gap of unknown length
* 43085 44226: contig of 1142 bp in length
* 44227 44326: gap of unknown length
* 44327 45730: contig of 1404 bp in length
* 45731 45830: gap of unknown length
* 45831 47585: contig of 1755 bp in length
* 47586 47685: gap of unknown length
* 47686 49636: contig of 1951 bp in length
* 49637 49736: gap of unknown length
* 49737 51465: contig of 1729 bp in length
* 51466 51565: gap of unknown length
* 51566 53094: contig of 1529 bp in length
* 53095 53194: gap of unknown length
* 53195 54598: contig of 1404 bp in length
* 54599 54698: gap of unknown length
* 54699 56537: contig of 1839 bp in length
* 56538 56637: gap of unknown length
* 56638 58197: contig of 1560 bp in length
* 58198 58297: gap of unknown length
* 58298 60205: contig of 1908 bp in length
* 60206 60305: gap of unknown length
* 60306 62035: contig of 1730 bp in length
* 62036 62135: gap of unknown length
* 62136 63583: contig of 1448 bp in length
* 63584 63683: gap of unknown length
* 63684 64712: contig of 1029 bp in length
* 64713 64812: gap of unknown length
* 64813 66180: contig of 1368 bp in length
* 66181 66280: gap of unknown length
* 66281 67973: contig of 1693 bp in length
* 67974 68073: gap of unknown length
* 68074 69216: contig of 1143 bp in length
* 69217 69316: gap of unknown length
* 69317 70320: contig of 1004 bp in length
* 70321 70420: gap of unknown length
* 70421 71895: contig of 1475 bp in length
* 71896 71995: gap of unknown length
* 71996 73282: contig of 1287 bp in length
* 73283 73382: gap of unknown length
* 73383 74885: contig of 1503 bp in length
* 74886 74985: gap of unknown length
* 74986 76544: contig of 1559 bp in length
* 76545 76644: gap of unknown length
* 76645 77867: contig of 1223 bp in length
* 77868 77967: gap of unknown length
* 77968 79384: contig of 1417 bp in length
* 79385 79484: gap of unknown length
* 79485 80930: contig of 1446 bp in length
* 80931 81030: gap of unknown length
* 81031 82098: contig of 1068 bp in length
* 82099 82198: gap of unknown length
* 82199 83570: contig of 1472 bp in length
* 83571 83770: gap of unknown length
* 83771 84871: contig of 1101 bp in length
* 84872 84971: gap of unknown length
* 84972 86513: contig of 1542 bp in length
* 86514 86613: gap of unknown length
* 86614 88046: contig of 1433 bp in length
* 88047 88146: gap of unknown length
* 88147 89321: contig of 1175 bp in length
* 89322 89421: gap of unknown length
* 89422 90439: contig of 1018 bp in length
* 90440 90539: gap of unknown length
* 90540 92219: contig of 1680 bp in length
* 92219 92319: gap of unknown length
* 92320 93428: contig of 1109 bp in length
* 93429 93528: gap of unknown length
* 93529 94590: contig of 1062 bp in length
* 94591 94690: gap of unknown length
* 94691 96150: contig of 1460 bp in length
* 96151 96250: gap of unknown length

```

```

* 96251 97757: contig of 1507 bp in length
* 97758 97857: gap of unknown length
* 97858 99164: contig of 1307 bp in length
* 99165 99264: gap of unknown length
* 99265 100715: contig of 1452 bp in length
* 100717 100816: gap of unknown length

Alignment Scores:
Pred. No.: 1.05e+03 Length: 107637
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

SEQ1-65TO79 (1-15) x AC094679 (1-107637)
QY 1 SerleuArgSerAlaHisIleuAla 8
Db 25714 AGTCACAGAGTGCACCTCGCC 25737

RESULT 33
HUMDGCRCEN HUMDGCRCEN 108400 bp DNA linear PRI 15-JUN-1996
LOCUS Homo sapiens DiGeorge syndrome critical region, centromeric end.
DEFINITION L77570
ACCESSION L77570
VERSION L77570.1 GI:1377755
KEYWORDS
SOURCE
ORGANISM Homo sapiens DNA.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 108400)
AUTHORS Gong, W., Emanuel, B.S., Collins, J., Kim, D.H., Wang, Z., Chen, F.,
Zhang, G., Roe, B., and Budarf, M.L.
TITLE A transcription map of the DiGeorge and velo-cardio-facial syndrome
JOURNAL minimal critical region on 22q11
MEDLINE Hum. Mol. Genet. 5 (6), 789-800 (1996)
PUBMED 96372815
FEATURES
PUBMED 8776594
Location/Qualifiers
source 1..108400
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 29924 a 28123 c 26278 g 24037 t 38 others
ORIGIN

Alignment Scores:
Pred. No.: 1.06e+03 Length: 108400
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 9 Gaps: 0

SEQ1-65TO79 (1-15) x HUMDGCRCEN (1-108400)
QY 1 SerleuArgSerAlaHisIleuAla 8
Db 61485 TCTTTGAGAGTGCACCTCGCC 61508

RESULT 34
AC128996 AC128996 108551 bp DNA linear HMG 02-AUG-2002
LOCUS Rattus norvegicus clone CH230-11M17, *** SEQUENCING IN PROGRESS
DEFINITION *** 62 unordered pieces.
ACCESSION AC128996
VERSION AC128996.2 GI:22038427
KEYWORDS HMG: HTGS PHASEL.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

```


REFERENCE
AUTHORS
Ratus.
1. (bases 1 to 108551)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C., Alshbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frawltz,P., Gabisi,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,C., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S., Moser,K., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonun,G., Ogasunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansay,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmali,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Welnsstock,G., and Gibbs,R.

TITLE
JOURNAL
Unpublished
2 (bases 1 to 108551)
Worley,K.C.

REFERENCE
AUTHORS
JOURNAL
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 108551)
Worley,K.C.

REFERENCE
AUTHORS
JOURNAL
Direct Submission
Submitted (02-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Aug 1, 2002 this sequence version replaced gi:21953916.

COMMENT
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GDRE
Center clone name: CH230-11M17
Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 50052 bases at least Q40
Consensus quality: 54741 bases at least Q30
Consensus quality: 57212 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1053: contig of 1053 bp in length
1054 1153: gap of unknown length
1154 2241: contig of 1088 bp in length
2242 2341: gap of unknown length
2342 3379: contig of 1038 bp in length
3380 3479: gap of unknown length
3480 4578: contig of 1099 bp in length
4579 5713: gap of unknown length
5714 5813: contig of 1035 bp in length
5814 7780: gap of unknown length
7781 7880: contig of 1967 bp in length
7881 8929: gap of unknown length
8930 9029: contig of 1049 bp in length
9030 10755: gap of unknown length
10756 10855: contig of 1726 bp in length
10856 11898: gap of unknown length
11899 11998: contig of 1043 bp in length
11999 13185: gap of unknown length
13186 13285: contig of 1187 bp in length
13286 14742: gap of unknown length
14743 14842: contig of 1457 bp in length
14843 15894: gap of unknown length
15894 15994: contig of 1055 bp in length
15994 17075: gap of unknown length
17075 17175: contig of 1081 bp in length
17176 18246: gap of unknown length
18247 18347: contig of 1071 bp in length
18347 19954: gap of unknown length
19955 20054: contig of 1608 bp in length
20054 21055: gap of unknown length
21055 21155: contig of 1001 bp in length
21156 22998: gap of unknown length
22999 23098: contig of 1843 bp in length
23099 24388: gap of unknown length
24389 24488: contig of 1290 bp in length
24489 25808: gap of unknown length
25809 25908: contig of 1320 bp in length
25909 26924: gap of unknown length
26924 27024: contig of 1016 bp in length
27024 27925: gap of unknown length
27925 28039: contig of 1015 bp in length
28040 28139: gap of unknown length
28140 29665: contig of 1526 bp in length
29666 31067: gap of unknown length
31067 31657: contig of 1302 bp in length
31658 32309: gap of unknown length
32310 32409: contig of 1142 bp in length
32410 33441: gap of unknown length
33441 33541: contig of 1032 bp in length
33542 34656: gap of unknown length
34656 34756: contig of 1115 bp in length
34757 35761: gap of unknown length
35762 35861: contig of 1005 bp in length
35862 37954: gap of unknown length
37955 38054: contig of 2093 bp in length
38055 39168: gap of unknown length
39169 41061: contig of 1114 bp in length
41062 41161: gap of unknown length
41162 42458: contig of 1793 bp in length
42459 42559: contig of 1297 bp in length
42559 43621: contig of 1063 bp in length

```

* 43622 43721: gap of unknown length
* 43722 45036: contig of 1315 bp in length
* 45037 45136: gap of unknown length
* 45137 46736: contig of 1600 bp in length
* 46737 46836: gap of unknown length
* 46837 48085: contig of 1249 bp in length
* 48086 48185: gap of unknown length
* 48186 49248: contig of 1063 bp in length
* 49249 49348: gap of unknown length
* 49349 50796: contig of 1448 bp in length
* 50797 50896: gap of unknown length
* 50897 52742: contig of 1846 bp in length
* 52743 52842: gap of unknown length
* 52843 54361: contig of 1519 bp in length
* 54362 54461: gap of unknown length
* 54462 56203: contig of 1742 bp in length
* 56204 56303: gap of unknown length
* 56304 57676: contig of 1373 bp in length
* 57677 57776: gap of unknown length
* 57777 59333: contig of 1557 bp in length
* 59334 59433: gap of unknown length
* 59434 60778: contig of 1345 bp in length
* 60779 60878: gap of unknown length
* 60879 62863: contig of 1985 bp in length
* 62864 62963: gap of unknown length
* 62964 64578: contig of 1615 bp in length
* 64579 64679: gap of unknown length
* 64679 66211: contig of 1533 bp in length
* 66212 66311: gap of unknown length
* 66312 68179: contig of 1868 bp in length
* 68180 68279: gap of unknown length
* 68280 70846: contig of 2567 bp in length
* 70847 70946: gap of unknown length
* 70947 73479: contig of 2533 bp in length
* 73480 73579: gap of unknown length
* 73580 74938: contig of 1359 bp in length
* 74939 75038: gap of unknown length
* 75039 76867: contig of 1829 bp in length
* 76868 76967: gap of unknown length
* 76968 78933: contig of 1966 bp in length
* 78934 79033: gap of unknown length
* 79034 80843: contig of 1810 bp in length
* 80844 80943: gap of unknown length
* 80944 83673: contig of 2730 bp in length

```

Alignment Scores:

```

Pred. No.: 1.06e+03 Length: 108551
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

```

SEQ1-65to79 (1-15) x AC128996 (1-108551)

```

Oy 1 SerLeuArgSerAlaHisLeuAla 8
Db 75594 AGCCTCAGAGTGCACCTCGCT 75617

RESULT 35
AC026736 AC026736 108661 bp DNA linear PRI 31-AUG-2000
LOCUS Homo sapiens chromosome 5 clone CTD-2350M8, complete sequence.
AC026736 AC026736
ACCESSION AC026736
VERSION AC026736.4 GI:9954680
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 108661)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 108661)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 108661)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Aug 31, 2000 this sequence version replaced gi:7711970.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99% of Sequence;
Estimated Total Number of Errors is 1.
Location/Qualifiers
1. 108661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2350M8"
BASE COUNT 30426 a 22057 c 22917 g 33261 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 1.06e+03 Length: 108661
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 9 Gaps: 0

```

SEQ1-65to79 (1-15) x AC026736 (1-108661)

```

Oy 1 SerLeuArgSerAlaHisLeuAla 8
Db 74759 TCCTCGGAGCGACATTTGGCA 74782

```

```

RESULT 36
AC109707 AC109707 109396 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-15701, *** SEQUENCING IN PROGRESS
AC109707 AC109707
ACCESSION AC109707.3 GI:21738035
VERSION AC109707.3
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 109396)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbaria,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,

```

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Lounsged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,
Massey,E., MaWhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwunu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivas,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 109396)
Morley,K.C.

Direct Submission
Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 109396)
Morley,K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18846951.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center Project name: GDBS
Center Clone name: CH230-15701
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 68820 bases at least Q40
Consensus quality: 73510 bases at least Q30
Consensus quality: 77739 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1152: contig of 1152 bp in length
* 1153 1252: gap of unknown length
* 1253 2704: contig of 1452 bp in length
* 2705 2804: gap of unknown length
* 2805 3912: contig of 1108 bp in length
* 3913 4012: gap of unknown length
* 4013 5034: contig of 1022 bp in length
* 5035 5134: gap of unknown length
* 5135 6205: contig of 1071 bp in length
* 6206 6305: gap of unknown length

6306 7385: contig of 1080 bp in length
* 7386 7485: gap of unknown length
* 7486 8677: contig of 1192 bp in length
* 8678 8777: gap of unknown length
* 8778 10177: contig of 1400 bp in length
* 10178 10277: gap of unknown length
* 10278 11788: contig of 1511 bp in length
* 11789 11888: gap of unknown length
* 11889 13480: contig of 1592 bp in length
* 13481 13580: gap of unknown length
* 13581 14600: contig of 1020 bp in length
* 14601 14700: gap of unknown length
* 14701 15891: contig of 1191 bp in length
* 15892 15991: gap of unknown length
* 15992 17208: contig of 1217 bp in length
* 17209 17308: gap of unknown length
* 17309 18887: contig of 1579 bp in length
* 18888 18987: gap of unknown length
* 18988 20995: contig of 2008 bp in length
* 20996 21095: gap of unknown length
* 21096 22918: contig of 1823 bp in length
* 22919 23018: gap of unknown length
* 23019 24262: contig of 1244 bp in length
* 24263 24362: gap of unknown length
* 24363 25621: contig of 1259 bp in length
* 25622 25721: gap of unknown length
* 25722 27130: contig of 1409 bp in length
* 27131 27230: gap of unknown length
* 27231 28460: contig of 1230 bp in length
* 28461 28560: gap of unknown length
* 28561 30254: contig of 1693 bp in length
* 30254 30353: gap of unknown length
* 30354 32124: contig of 1771 bp in length
* 32125 32224: gap of unknown length
* 32225 33689: contig of 1465 bp in length
* 33690 33789: gap of unknown length
* 33790 35176: contig of 1387 bp in length
* 35177 35276: gap of unknown length
* 35277 37357: contig of 2081 bp in length
* 37358 37457: gap of unknown length
* 37458 39566: contig of 2109 bp in length
* 39567 39666: gap of unknown length
* 39667 41918: contig of 2252 bp in length
* 41919 42018: gap of unknown length
* 42019 43673: contig of 1655 bp in length
* 43674 43773: gap of unknown length
* 43774 45144: contig of 1371 bp in length
* 45145 45244: gap of unknown length
* 45245 47961: contig of 2717 bp in length
* 47962 48061: gap of unknown length
* 48062 49681: contig of 1620 bp in length
* 49682 49781: gap of unknown length
* 49782 51802: contig of 2021 bp in length
* 51803 51902: gap of unknown length
* 51903 53819: contig of 1917 bp in length
* 53820 53919: gap of unknown length
* 53920 55692: contig of 1773 bp in length
* 55693 55792: gap of unknown length
* 55793 57334: contig of 1542 bp in length
* 57335 57434: gap of unknown length
* 57435 58665: contig of 1231 bp in length
* 58666 58765: gap of unknown length
* 58766 60001: contig of 1236 bp in length
* 60002 60101: gap of unknown length
* 60102 62593: contig of 2452 bp in length
* 62594 62693: gap of unknown length
* 62694 64737: contig of 2044 bp in length
* 64738 64837: gap of unknown length
* 64838 67007: contig of 2170 bp in length
* 67008 67107: gap of unknown length
* 67108 68376: contig of 1269 bp in length
* 68377 68476: gap of unknown length
* 68477 70582: contig of 2106 bp in length

```

* 70583 70682: gap of unknown length
* 70683 72254: contig of 1572 bp in length
* 72255 72354: gap of unknown length
* 72355 74323: contig of 1969 bp in length
* 74324 74423: gap of unknown length
* 74424 76330: contig of 1907 bp in length
* 76331 76430: gap of unknown length
* 76431 79126: contig of 2696 bp in length
* 79127 79226: gap of unknown length
* 79227 81807: contig of 2581 bp in length
* 81808 81907: gap of unknown length
* 81908 84447: contig of 2540 bp in length
* 84448 84547: gap of unknown length
* 84548 88024: contig of 3477 bp in length
* 88025 88124: gap of unknown length
* 88125 92389: contig of 4165 bp in length
* 92390 92390: gap of unknown length
* 92390 98425: contig of 6036 bp in length
* 98426 98525: gap of unknown length
* 98526 103173: contig of 4648 bp in length
* 103174 103273: gap of unknown length
* 103274 109396: contig of 6123 bp in length.

FEATURES
  source          1..109396
                  Location/Qualifiers

Alignment Scores:
Pred. No.:      1.07e+03      Length:      109396
Score:          8.00          Matches:      8
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    53.33%        Indels:        0
DB:              2            Gaps:            0

SEQ1-65to79 (1-15) x AC109707 (1-109396)
QY      1 SerLeuArgSerAlaHisLeuAla 8
Db 76509 AGCCTTAGAGATGCCACCTGCC 76532

RESULT 37
AC128969
LOCUS      AC128969      114789 bp      DNA      linear      HTG 24-JUL-2002
DEFINITION Rattus norvegicus clone CH230-283G2, *** SEQUENCING IN PROGRESS
ACCESSION AC128969
VERSION   AC128969.1 GI:21953761
KEYWORDS  HTG; HTGS_PHASE1.
SOURCE    Rattus norvegicus.
ORGANISM  Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
          1 (bases 1 to 114789)
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbarta,J., Benton,J., Blinag,K., Blankenburg,K., Bonnin,D.,
Bouck,D., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,
Homs,J.F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W., Lounsged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meadow,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunnu,G.,
Oragunye,N., Oriedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Qules,M., Ren,Y.,
Rives,M., Rojas,A., Rojuboan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 114789)
Worley,K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: KBXI
Center clone name: CH230-283G2
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990328
Consensus quality: 78414 bases at least Q40
Consensus quality: 82108 bases at least Q30
Consensus quality: 85781 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1025: contig of 1025 bp in length
1026 1125: gap of unknown length
1126 2579: contig of 1454 bp in length
2580 2679: gap of unknown length
2680 3848: contig of 1169 bp in length
3849 3948: gap of unknown length
3949 5207: contig of 1259 bp in length
5208 5308: gap of unknown length
5309 6740: contig of 1433 bp in length
6741 6840: gap of unknown length
6841 7981: contig of 1141 bp in length
7982 8081: gap of unknown length
8082 9871: contig of 1799 bp in length
9872 9971: gap of unknown length
9972 11427: contig of 1455 bp in length
11428 11526: gap of unknown length
11527 12609: contig of 1083 bp in length
12610 12710: gap of unknown length
12711 13885: contig of 1176 bp in length

```

* 13886 13985: gap of unknown length
* 13986 15274: contig of 1289 bp in length
* 15275 15374: gap of unknown length
* 15375 17715: contig of 2341 bp in length
* 17716 17815: gap of unknown length
* 17816 19068: contig of 1253 bp in length
* 19069 19168: gap of unknown length
* 19169 20484: contig of 1316 bp in length
* 20485 20584: gap of unknown length
* 20585 22254: contig of 1670 bp in length
* 22255 22354: gap of unknown length
* 22355 23426: contig of 1072 bp in length
* 23427 23526: gap of unknown length
* 23527 24921: contig of 1395 bp in length
* 24922 25021: gap of unknown length
* 25022 26194: contig of 1173 bp in length
* 26195 26294: gap of unknown length
* 26295 27482: contig of 1188 bp in length
* 27483 27582: gap of unknown length
* 27583 29451: contig of 1869 bp in length
* 29452 31325: gap of unknown length
* 31326 31425: contig of 1774 bp in length
* 31426 32712: gap of unknown length
* 32713 32812: contig of 1287 bp in length
* 32813 34615: contig of 1803 bp in length
* 34616 34715: gap of unknown length
* 34716 36615: contig of 1900 bp in length
* 36616 36715: gap of unknown length
* 36716 38473: contig of 1758 bp in length
* 38474 38573: gap of unknown length
* 38574 40605: contig of 2032 bp in length
* 40606 42853: gap of unknown length
* 42854 42953: contig of 2148 bp in length
* 42954 45191: gap of unknown length
* 45192 45291: contig of 2238 bp in length
* 45292 47147: gap of unknown length
* 47148 47247: contig of 1856 bp in length
* 47248 48910: gap of unknown length
* 48911 49010: contig of 1663 bp in length
* 49011 50703: gap of unknown length
* 50704 50803: contig of 1693 bp in length
* 50804 52366: gap of unknown length
* 52367 52466: contig of 1563 bp in length
* 52467 54624: gap of unknown length
* 54625 54724: contig of 2158 bp in length
* 54725 57204: gap of unknown length
* 57205 57304: contig of 2480 bp in length
* 57305 59915: gap of unknown length
* 59916 60015: contig of 2611 bp in length
* 60016 63063: gap of unknown length
* 63064 65786: contig of 3048 bp in length
* 65787 65886: gap of unknown length
* 65887 68161: contig of 2623 bp in length
* 68162 68261: gap of unknown length
* 68262 71290: contig of 2275 bp in length
* 71291 71390: gap of unknown length
* 71391 73552: contig of 3029 bp in length
* 73553 73652: gap of unknown length
* 73653 75936: contig of 2162 bp in length
* 75937 76036: gap of unknown length
* 76037 79206: contig of 2284 bp in length
* 79207 79306: gap of unknown length
* 79307 82322: contig of 3170 bp in length
* 82323 82422: gap of unknown length
* 82423 86777: contig of 3016 bp in length
* 86778 86877: gap of unknown length
* 86879 90387: contig of 4355 bp in length
* 90388 90487: gap of unknown length
* 90488 94881: contig of 3510 bp in length
* 94882 94981: gap of unknown length
* 94982 99220: contig of 4239 bp in length
* 99221 99320: gap of unknown length
* 99321 104222: contig of 4902 bp in length
* 104223 104323: gap of unknown length
* 104323 109517: contig of 5195 bp in length
* 109518 109617: gap of unknown length
* 109618 114789: contig of 5172 bp in length.
Location/Qualifiers
1. 114789
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-283G2"

BASE COUNT 30903 a 22048 c 22499 g 31310 t 8029 others
ORIGIN

Alignment Scores:

Pred. No.: 1.11e+03 Length: 114789
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

SEQ1-65TO79 (1-15) x AC128969 (1-114789)

Oy 1 SerLeuArgSerAlaHisLeuAla 8

Db 27325 AGCCTCAGAGTGGCCACCTCGCA 27348

RESULT 38

AC131517

LOCUS

DEFINITION

AC131517

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 118681)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Giller,R., Gottell,J.H., Gueraa,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichatage,O., Lien,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwoung,G., Otagunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Qilles,M., Ren,Y.,

Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telitrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 118681)
Worley,K.C.

Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc.helpebcm.tmc.edu
----- Project Information
Center project name: T0M2
Center clone name: CH230-2B9
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 29641 bases at least Q40
Consensus quality: 34185 bases at least Q30
Consensus quality: 38055 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/Gendank_draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1031: contig of 1031 bp in length
* 1032 1131: gap of unknown length
* 1132 2131: contig of 1000 bp in length
* 2132 2231: gap of unknown length
* 2232 3350: contig of 1119 bp in length
* 3351 3450: gap of unknown length
* 3451 4574: contig of 1124 bp in length
* 4575 4674: gap of unknown length
* 4675 6084: contig of 1410 bp in length
* 6085 6184: gap of unknown length
* 6185 7186: contig of 1002 bp in length
* 7187 7286: gap of unknown length
* 7287 8833: contig of 1547 bp in length
* 8834 8934: gap of unknown length
* 8934 10405: contig of 1472 bp in length
* 10406 10505: gap of unknown length
* 10506 11730: contig of 1225 bp in length
* 11731 11830: gap of unknown length
* 11831 13314: contig of 1484 bp in length
* 13315 13414: gap of unknown length
* 13415 14786: contig of 1372 bp in length
* 14787 14886: gap of unknown length
* 14887 16364: contig of 1478 bp in length
* 16365 16464: gap of unknown length
* 16465 17790: contig of 1326 bp in length
* 17791 17890: gap of unknown length
* 17891 19337: contig of 1447 bp in length
* 19338 19437: gap of unknown length
* 19438 20807: contig of 1370 bp in length

20808 20907: gap of unknown length
* 20908 22043: contig of 1136 bp in length
* 22044 22143: gap of unknown length
* 22144 23184: contig of 1041 bp in length
* 23185 23284: gap of unknown length
* 23285 24851: contig of 1567 bp in length
* 24852 24951: gap of unknown length
* 24952 26209: contig of 1258 bp in length
* 26210 26309: gap of unknown length
* 26310 27761: contig of 1452 bp in length
* 27762 27861: gap of unknown length
* 27862 29206: contig of 1345 bp in length
* 29207 29306: gap of unknown length
* 29307 30342: contig of 1036 bp in length
* 30343 30442: gap of unknown length
* 30443 31975: contig of 1534 bp in length
* 31977 32077: gap of unknown length
* 32077 33855: contig of 1779 bp in length
* 33856 33955: gap of unknown length
* 33956 35340: contig of 1385 bp in length
* 35341 35440: gap of unknown length
* 35441 36720: contig of 1280 bp in length
* 36721 36820: gap of unknown length
* 36821 38421: contig of 1601 bp in length
* 38422 38521: gap of unknown length
* 38522 39534: contig of 1013 bp in length
* 39535 39634: gap of unknown length
* 39635 41249: contig of 1615 bp in length
* 41250 41349: gap of unknown length
* 41350 43034: contig of 1685 bp in length
* 43035 43134: gap of unknown length
* 43135 44233: contig of 1099 bp in length
* 44234 44333: gap of unknown length
* 44334 46065: contig of 1732 bp in length
* 46066 46165: gap of unknown length
* 46166 47833: contig of 1668 bp in length
* 47834 47933: gap of unknown length
* 47934 49208: contig of 1275 bp in length
* 49209 49308: gap of unknown length
* 49309 50760: contig of 1452 bp in length
* 50761 50860: gap of unknown length
* 50861 52682: contig of 1822 bp in length
* 52683 52782: gap of unknown length
* 52783 54216: contig of 1434 bp in length
* 54217 54316: gap of unknown length
* 54317 55449: contig of 1133 bp in length
* 55450 55548: gap of unknown length
* 55550 57308: contig of 1759 bp in length
* 57309 57408: gap of unknown length
* 57409 59193: contig of 1785 bp in length
* 59194 59293: gap of unknown length
* 59294 60574: contig of 1281 bp in length
* 60575 60674: gap of unknown length
* 60675 62172: contig of 1498 bp in length
* 62173 62272: gap of unknown length
* 62273 63401: contig of 1129 bp in length
* 63402 63501: gap of unknown length
* 63502 65247: contig of 1746 bp in length
* 65248 65347: gap of unknown length
* 65348 67360: contig of 2013 bp in length
* 67361 67460: gap of unknown length
* 67461 69390: contig of 1930 bp in length
* 69391 69490: gap of unknown length
* 69491 71020: contig of 1530 bp in length
* 71021 71120: gap of unknown length
* 71121 72335: contig of 1215 bp in length
* 72336 72435: gap of unknown length
* 72436 72445: contig of 1810 bp in length
* 72446 74345: gap of unknown length
* 74346 76137: contig of 1792 bp in length
* 76138 76237: gap of unknown length
* 76238 78535: contig of 2298 bp in length
* 78536 78635: gap of unknown length

```

* 78636 79847: contig of 1212 bp in length
* 79848 79947: gap of unknown length
* 79948 82456: contig of 2509 bp in length
* 82456 82557: gap of unknown length
* 82557 84509: contig of 1953 bp in length
* 84510 84609: gap of unknown length
* 84610 86889: contig of 2280 bp in length
* 86889 86989: gap of unknown length
* 86989 88505: contig of 1516 bp in length
* 88505 88606: gap of unknown length
* 88606 89850: contig of 1245 bp in length
* 89850 89950: gap of unknown length

```

Alignment Scores:

```

Pred. No.: 1.15e+03 Length: 118681
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

```

SEQ1-65TO79 (1-15) x AC131517 (1-118681)

```

OY 1 SerleuArgSerAlaHisLeuAla 8
Db 1616 AGCCTCAGAGTGCCTGCTGCC 1639

```

RESULT 39

```

AC119340 AC119340 119785 bp DNA linear HTG 18-JUL-2002
LOCUS Rattus norvegicus clone CH230-354C16, *** SEQUENCING IN PROGRESS
DEFINITION

```

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

```

AC119340 AC119340 119785 bp DNA linear HTG 18-JUL-2002
AC119340 AC119340 3 GI:21746752
HTG, HTGS_PHASE1.
Rattus norvegicus
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 119785)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaral-Lungue,H.C., Are,J.R., Ayale,M., Banks,T.,
Barberia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowls,B., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsdon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Moser,G., Miner,Z., Mitchell,T., Mohabadi,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oyler,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojucokan,I., Rolfe,M., Ruiz,S., Severy,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,

```

COMMENT

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

COMMENT

```

***** Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc.help@bcm.tmc.edu
***** Project Information
Center project name: GUVS
Center clone name: CH230-354C16
***** Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 91219 bases at least Q40
Consensus quality: 94090 bases at least Q30
Consensus quality: 95811 bases at least Q20
*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1059: contig of 1059 bp in length
* 1060 1159: gap of unknown length
* 1160 2166: contig of 1007 bp in length
* 2167 2266: gap of unknown length
* 2267 3777: contig of 1511 bp in length
* 3778 3877: gap of unknown length
* 3878 4993: contig of 1116 bp in length
* 4994 5093: gap of unknown length
* 5094 6168: contig of 1075 bp in length
* 6169 6268: gap of unknown length
* 6269 7326: contig of 1058 bp in length
* 7327 7426: gap of unknown length
* 7427 8836: contig of 1410 bp in length
* 8837 8936: gap of unknown length
* 8937 10047: contig of 1111 bp in length
* 10048 10147: gap of unknown length
* 10148 11931: contig of 1784 bp in length
* 11932 12031: gap of unknown length
* 12032 13856: contig of 1825 bp in length
* 13857 13956: gap of unknown length
* 13957 16221: contig of 2265 bp in length
* 16222 16321: gap of unknown length
* 16322 17533: contig of 1212 bp in length
* 17534 17633: gap of unknown length

```

```

* 17634 18797: contig of 1164 bp in length
* 18798 18897: gap of unknown length
* 18898 20460: contig of 1563 bp in length
* 20461 20560: gap of unknown length
* 20561 22303: contig of 2370 bp in length
* 22303 23030: gap of unknown length
* 23031 26397: contig of 3367 bp in length
* 26398 26497: gap of unknown length
* 26498 28790: contig of 2293 bp in length
* 28791 28890: gap of unknown length
* 28891 32306: contig of 3416 bp in length
* 32307 32406: gap of unknown length
* 32407 35895: contig of 3489 bp in length
* 35896 35995: gap of unknown length
* 35996 39348: contig of 3353 bp in length
* 39349 39448: gap of unknown length
* 39449 42377: contig of 2929 bp in length
* 42378 42477: gap of unknown length
* 42478 45875: contig of 3398 bp in length
* 45876 45975: gap of unknown length
* 45976 50586: contig of 4611 bp in length
* 50587 50686: gap of unknown length
* 50687 54403: contig of 3717 bp in length
* 54404 54503: gap of unknown length
* 54504 59448: contig of 4945 bp in length
* 59449 59548: gap of unknown length
* 59549 63988: contig of 4440 bp in length
* 63989 64088: gap of unknown length
* 64089 68524: contig of 4436 bp in length
* 68525 68624: gap of unknown length
* 68625 73152: contig of 4528 bp in length
* 73153 73252: gap of unknown length
* 73253 78551: contig of 5299 bp in length
* 78552 78651: gap of unknown length
* 78652 83501: contig of 4850 bp in length
* 83502 83601: gap of unknown length
* 83602 87143: contig of 3542 bp in length
* 87144 87243: gap of unknown length
* 87244 93854: contig of 6611 bp in length
* 93855 93955: gap of unknown length
* 93956 102139: contig of 8185 bp in length
* 102140 102239: gap of unknown length
* 102240 111853: contig of 9614 bp in length
* 111854 111953: gap of unknown length
* 111954 1119785: contig of 7832 bp in length.

```

```

FEATURES
  source      1..119785
              /organism="Rattus norvegicus"
              /db_xref="taxon:10116"
              /clone="CH230-354C16"
BASE COUNT   35815 a 22173 c 22058 g 36309 t 3430 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1.16e+03      Length:      119785
Score:          8.00          Matches:      8
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:     53.33%       Indels:        0
DB:              2           Gaps:          0

```

SEQ1-65TO79 (1-15) x AC119340 (1-119785)

OY 1 SerleuArgSerAlaHisIleuAla 8
 |||
 Db 14225 AGCCTTAGGAGTGCCACCTCGCT 14202

```

RESULT 40      AC120314      121080 bp  DNA      linear  HTG 18-JUL-2002
LOCUS          Rattus norvegicus clone CH230-402E14, **** SEQUENCING IN PROGRESS
DEFINITION     *** 35 unordered pieces.
ACCESSION      AC120314

```

```

VERSION
AC120314.2 GI:21747275
HTG: HTGS PHASE1.
KEYWORDS
Norway rat.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 121080)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devalla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,J.F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,B., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzler,M.,
Miner,G., Miner,Z., Mitchell,T., Monabbat,K., Morgan,M., Morris,S.,
Moser,M., Neel,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenwo,S., Ogih,M., Okumura,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherrer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,
Soedergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
Sulton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 121080)
Worley,K.C.
Direct Submission
Submitted (06-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 121080)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20454688.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYVP
Center clone name: CH230-402E14
----- Summary Statistics
Sequencing vector: Plasmid;

```


Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 9212 bases at least Q40
Consensus quality: 95963 bases at least Q30
Consensus quality: 98562 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1015: contig of 1015 bp in length
1016 1115: gap of unknown length
1116 2318: contig of 1203 bp in length
2319 2418: gap of unknown length
2419 3998: contig of 1580 bp in length
3999 4099: gap of unknown length
4099 5609: contig of 1511 bp in length
5610 5709: gap of unknown length
5710 7164: contig of 1455 bp in length
7165 7264: gap of unknown length
7265 8388: contig of 1124 bp in length
8389 8489: gap of unknown length
8489 10393: contig of 1905 bp in length
10394 10493: gap of unknown length
10494 11814: contig of 1321 bp in length
11815 11914: gap of unknown length
11915 14309: contig of 2395 bp in length
14310 14409: gap of unknown length
14410 16493: contig of 2084 bp in length
16494 16593: gap of unknown length
16594 19202: contig of 2609 bp in length
19203 19302: gap of unknown length
19303 22300: contig of 2998 bp in length
22301 22401: gap of unknown length
22401 24919: contig of 2519 bp in length
24920 25019: gap of unknown length
25019 26638: contig of 1619 bp in length
26639 26738: gap of unknown length
26739 29220: contig of 2482 bp in length
29221 29320: gap of unknown length
29321 33130: contig of 3810 bp in length
33131 33230: gap of unknown length
33231 35377: contig of 2147 bp in length
35378 35477: gap of unknown length
35478 38105: contig of 2628 bp in length
38106 38205: gap of unknown length
38206 41333: contig of 3128 bp in length
41334 41433: gap of unknown length
41434 43740: contig of 2307 bp in length
43741 43841: gap of unknown length
43841 45966: contig of 2126 bp in length
45967 46066: gap of unknown length
46067 50108: contig of 4042 bp in length
50109 50208: gap of unknown length
50209 53272: contig of 3064 bp in length
53273 53372: gap of unknown length
53373 56127: contig of 2755 bp in length
56128 56227: gap of unknown length
56228 62222: contig of 5995 bp in length
62223 62322: gap of unknown length
62323 65871: gap of unknown length
65872 70954: contig of 4983 bp in length
70955 71054: gap of unknown length
71055 77892: contig of 6838 bp in length
77893 82120: contig of 4128 bp in length

82121 82220: gap of unknown length
82221 87930: contig of 5710 bp in length
87931 88030: gap of unknown length
88031 94562: contig of 6532 bp in length
94563 94662: gap of unknown length
94663 99940: contig of 5278 bp in length
99941 100040: gap of unknown length
100041 108790: contig of 8750 bp in length
108791 108891: gap of unknown length
108891 114163: contig of 5273 bp in length
114164 114263: gap of unknown length
114264 121080: contig of 6817 bp in length.
Location/Qualifiers
1. .121080
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-402E14"

BASE COUNT 34478 a 23795 c 22782 g 35719 t 4306 others
ORIGIN

Alignment Scores:
Pred. No.: 1.17e+03 Length: 121080
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

SEQ1-65TO79 (1-15) x AC120314 (1-121080)

Qy 1 SerleuArgSerAlaHisLeuAla 8
Db 71529 AGCCTCAGCAGTCCACCCCTCGCT 71552

RESULT 41
AC091350/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-1B16, *** SEQUENCING IN PROGRESS ***,
62 unordered pieces.
ACCESSION
AC091350 GI:22474745
VERSION
AC091350.5
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 122517)
REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barberia,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carrion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.U., Draper,H., Dugan-Rocha,S., Durin,H.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hatt,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsif,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., Kling,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,Z., Lichtarge,O., Liew,C., Liu,J., Liu,W., Louised,H.,
Lozao,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Moser, N., Nickerson, E., Nwankwo, S., Oguh, M., Okunolu, G., Oranuy, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pichens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sotak, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 122517)
Worley, K.C.
Direct Submission
Submitted (18-APR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 122517)
Worley, K.C.
Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Aug 24, 2002 this sequence version replaced gi:21953961.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUEA
Center clone name: CH230-1B16
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-primer Bodypy: 58% of reads
Chemistry: Dye-terminator Big Dye: 42% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 70875 bases at least Q40
Consensus quality: 76728 bases at least Q30
Consensus quality: 79150 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1168: contig of 1168 bp in length
* 1169 1268: gap of unknown length
* 1269 1275: contig of 1007 bp in length
* 1276 2375: gap of unknown length
* 2376 3438: contig of 1063 bp in length
* 3439 3538: gap of unknown length
* 3539 4613: contig of 1075 bp in length
* 4614 4714: gap of unknown length
* 4714 4715: gap of unknown length
* 4715 5745: contig of 1032 bp in length
* 5746 5845: gap of unknown length
* 5846 6907: contig of 1062 bp in length
* 6908 7007: gap of unknown length
* 7008 8039: contig of 1032 bp in length
* 8040 8139: gap of unknown length
* 8140 9451: contig of 1312 bp in length

9452 9551: gap of unknown length
* 9552 10914: contig of 1363 bp in length
* 10915 11014: gap of unknown length
* 11015 12260: contig of 1246 bp in length
* 12261 12360: gap of unknown length
* 12361 13515: contig of 1155 bp in length
* 13516 13615: gap of unknown length
* 13616 14668: contig of 1053 bp in length
* 14669 14768: gap of unknown length
* 14769 16180: contig of 1412 bp in length
* 16181 16280: gap of unknown length
* 16281 17589: contig of 1309 bp in length
* 17590 17689: gap of unknown length
* 17690 18836: contig of 1147 bp in length
* 18837 18936: gap of unknown length
* 18937 20190: contig of 1254 bp in length
* 20191 20290: gap of unknown length
* 20291 21840: contig of 1550 bp in length
* 21841 21940: gap of unknown length
* 21941 23202: contig of 1262 bp in length
* 23203 23303: gap of unknown length
* 23303 24362: contig of 1060 bp in length
* 24363 24462: gap of unknown length
* 24463 25895: contig of 1433 bp in length
* 25896 25995: gap of unknown length
* 25996 27177: contig of 1182 bp in length
* 27178 27277: gap of unknown length
* 27278 28596: contig of 1419 bp in length
* 28597 28796: gap of unknown length
* 28797 30378: contig of 1582 bp in length
* 30379 30478: gap of unknown length
* 30479 31855: contig of 1377 bp in length
* 31856 31955: gap of unknown length
* 31956 33142: contig of 1186 bp in length
* 33142 33242: gap of unknown length
* 33242 34423: contig of 1182 bp in length
* 34423 34523: gap of unknown length
* 34524 3638: contig of 2115 bp in length
* 3639 36738: gap of unknown length
* 36739 37784: contig of 1046 bp in length
* 37785 37884: gap of unknown length
* 37885 39305: contig of 1421 bp in length
* 39306 39405: gap of unknown length
* 39406 40903: contig of 1498 bp in length
* 40904 41003: gap of unknown length
* 41004 42118: contig of 1115 bp in length
* 42119 42218: gap of unknown length
* 42219 43416: contig of 1198 bp in length
* 43417 43516: gap of unknown length
* 43517 45350: contig of 1834 bp in length
* 45351 45450: gap of unknown length
* 45451 47159: contig of 1709 bp in length
* 47160 47259: gap of unknown length
* 47260 48641: contig of 1382 bp in length
* 48642 48741: gap of unknown length
* 48742 50087: contig of 1346 bp in length
* 50088 50187: gap of unknown length
* 50188 52135: contig of 1948 bp in length
* 52136 52235: gap of unknown length
* 52236 54710: contig of 2475 bp in length
* 54711 54810: gap of unknown length
* 54811 55897: contig of 1087 bp in length
* 55898 55997: gap of unknown length
* 55998 57628: contig of 1631 bp in length
* 57629 57728: gap of unknown length
* 57729 59568: contig of 1840 bp in length
* 59569 59668: gap of unknown length
* 59669 62337: contig of 2668 bp in length
* 62337 62436: gap of unknown length
* 62437 64086: contig of 1650 bp in length
* 64087 64186: gap of unknown length
* 64187 66576: contig of 2390 bp in length
* 66577 66676: gap of unknown length

```

* 66677 68642: contig of 1966 bp in length
* 68643 68742: gap of unknown length
* 68743 70703: contig of 1961 bp in length
* 70704 70803: gap of unknown length
* 70804 73370: contig of 2367 bp in length
* 73371 73470: gap of unknown length
* 73471 75651: contig of 2181 bp in length
* 75652 75751: gap of unknown length
* 75752 78204: contig of 2353 bp in length
* 78205 81006: contig of 2802 bp in length
* 81007 81106: gap of unknown length
* 81107 83508: contig of 2402 bp in length
* 83509 83608: gap of unknown length
* 83609 86378: contig of 2770 bp in length
* 86379 86478: gap of unknown length
* 86479 88803: contig of 2325 bp in length

```

Alignment Scores:

```

Pred. No.: 1.18e+03 Length: 122517
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0

```

SEQ1-65TO79 (1-15) x AC091350 (1-122517)

OY 1 SerleuArgSerAlaHisLeuAla 8

Db 33860 AGCCTCAGAGTGCCTCCTGCCC 33837

RESULT 42

AC103101/c 124579 bp DNA linear HTG 13-JUL-2002

LOCUS Rattus norvegicus clone CH230-47P11, *** SEQUENCING IN PROGRESS

DEFINITION *** 84 unordered pieces.

AC103101.3 GI:21731028

VERSION HTG: HTCS_PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Rattus.

1 (bases 1 to 124579)

Muzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbieri,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,

Bouch,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dalhorne,S.R., David,R.,

DeValla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

```

Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Plickens,R., Prime,A., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojuben,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vasquez,L., Vera,V., Villalón,D., Vlnson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 124579)
Worley,K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 124579)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17974587.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJFC
Center clone name: CH230-47P11
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 23868 bases at least Q40
Consensus quality: 26331 bases at least Q30
Consensus quality: 28175 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 84 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1067: contig of 1067 bp in length
* 1068 1167: gap of unknown length
* 1168 2324: contig of 1157 bp in length
* 2325 2424: gap of unknown length
* 2425 3973: contig of 1549 bp in length
* 3974 4073: gap of unknown length
* 4074 5419: contig of 1346 bp in length
* 5420 5519: gap of unknown length
* 5520 7023: contig of 1504 bp in length
* 7024 7123: gap of unknown length
* 7124 8344: contig of 1221 bp in length
* 8345 8444: gap of unknown length
* 8445 9485: contig of 1041 bp in length
* 9486 9585: gap of unknown length
* 9586 10610: contig of 1025 bp in length
* 10611 10710: gap of unknown length
* 10711 11939: contig of 1229 bp in length
* 11940 12039: gap of unknown length
* 12040 13197: contig of 1158 bp in length

```

```

* 13198 13297: gap of unknown length
* 13298 14442: contig of 1145 bp in length
* 14443 14542: gap of unknown length
* 14543 15627: contig of 1085 bp in length
* 15628 15727: gap of unknown length
* 15728 16887: contig of 1160 bp in length
* 16888 16987: gap of unknown length
* 16988 18016: contig of 1029 bp in length
* 18017 18116: gap of unknown length
* 18117 19126: contig of 1010 bp in length
* 19127 19226: gap of unknown length
* 19227 20788: contig of 1562 bp in length
* 20789 20888: gap of unknown length
* 20889 22135: contig of 1247 bp in length
* 22136 22235: gap of unknown length
* 22236 23992: contig of 1757 bp in length
* 23993 24092: gap of unknown length
* 24093 25671: contig of 1579 bp in length
* 25672 25771: gap of unknown length
* 25772 26834: contig of 1063 bp in length
* 26835 26934: gap of unknown length
* 26935 27984: contig of 1050 bp in length
* 27985 28084: gap of unknown length
* 28085 29126: contig of 1042 bp in length
* 29127 29226: gap of unknown length
* 29227 30244: contig of 1018 bp in length
* 30245 30344: gap of unknown length
* 30345 31372: contig of 1028 bp in length
* 31373 31473: gap of unknown length
* 31473 32565: contig of 1093 bp in length
* 32566 32665: gap of unknown length
* 32666 33726: contig of 1061 bp in length
* 33727 33826: gap of unknown length
* 33827 35562: contig of 1736 bp in length
* 35563 35662: gap of unknown length
* 35663 37113: contig of 1451 bp in length
* 37114 37213: gap of unknown length
* 37214 38449: contig of 1236 bp in length
* 38450 38549: gap of unknown length
* 38550 39613: contig of 1064 bp in length
* 39614 39713: gap of unknown length
* 39714 41384: contig of 1671 bp in length
* 41385 41484: gap of unknown length
* 41485 42566: contig of 1082 bp in length
* 42567 42666: gap of unknown length
* 42667 43825: contig of 1159 bp in length
* 43826 43925: gap of unknown length
* 43926 45020: contig of 1095 bp in length
* 45021 45120: gap of unknown length
* 45121 46219: contig of 1099 bp in length
* 46220 46319: gap of unknown length
* 46320 47606: contig of 1287 bp in length
* 47607 47706: gap of unknown length
* 47707 48919: contig of 1213 bp in length
* 48920 49019: gap of unknown length
* 49020 50385: contig of 1366 bp in length
* 50386 50485: gap of unknown length
* 50486 51777: contig of 1292 bp in length
* 51778 51877: gap of unknown length
* 51878 53545: contig of 1668 bp in length
* 53546 53645: gap of unknown length
* 53646 55151: contig of 1506 bp in length
* 55152 55251: gap of unknown length
* 55252 56519: contig of 1268 bp in length
* 56520 57857: gap of unknown length
* 57858 57957: contig of 1238 bp in length
* 57959 59032: gap of unknown length
* 59033 59132: contig of 1075 bp in length
* 59133 60672: gap of unknown length
* 60673 60772: gap of unknown length
* 60773 61823: contig of 1051 bp in length
* 61824 61923: gap of unknown length

```

```

* 61924 63420: contig of 1497 bp in length
* 63421 63520: gap of unknown length
* 63521 65509: contig of 1989 bp in length
* 65510 65609: gap of unknown length
* 65610 66640: contig of 1031 bp in length
* 66641 66740: gap of unknown length
* 66741 68005: contig of 1265 bp in length
* 68006 68105: gap of unknown length
* 68106 69745: contig of 1641 bp in length
* 69747 69846: gap of unknown length
* 69847 71056: contig of 1210 bp in length
* 71057 71156: gap of unknown length
* 71157 72367: contig of 1211 bp in length
* 72368 72467: gap of unknown length
* 72468 73594: contig of 1127 bp in length

Alignment Scores:
Pred. No.: 1.2e+03 Length: 124579
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
Db: 2 Gaps: 0

SEQ1-657079 (1-15) x AC103101 (1-124579)
Oy 1 SerLeuArgSerAlaHisLeuAla 8
Db 97943 AGCCTCAGAGTGCCACCTGCGC 97920

RESULT 43
AC124138
LOCUS
DEFINITION
Rattus norvegicus clone CH230-392N11, *** SEQUENCING IN PROGRESS
AC124138
VERSION
AC124138.1 GI:21362133
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 126809)
REFERENCE
AUTHORS
Aisbrooks,S.L., Amaralunga,H.C., Aze,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louisedge,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,B.,
Massey,E., Mawliny,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

```

Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Severe, G.,
 Scherer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I.,
 Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-More, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Unpublished
 Direct Submission
 2 (bases 1 to 126809)

Worley, K.C.
 Direct Submission
 Submitted (10-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 126809)

Worley, K.C.
 Direct Submission
 Submitted (21-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: KAEI
 Center clone name: CH230-392N11
 ----- Summary Statistics -----
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 77471 bases at least Q40
 Consensus quality: 80999 bases at least Q30
 Consensus quality: 83473 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1009: contig of 1009 bp in length
 * 1010 1109: gap of unknown length
 * 1110 2200: contig of 1091 bp in length
 * 2201 2300: gap of unknown length
 * 2301 3337: contig of 1037 bp in length
 * 3338 3437: gap of unknown length
 * 3438 5001: contig of 1564 bp in length
 * 5002 5101: gap of unknown length
 * 5102 6599: contig of 1498 bp in length
 * 6600 6699: gap of unknown length
 * 6700 7901: contig of 1202 bp in length
 * 7902 8001: gap of unknown length
 * 8002 9051: contig of 1050 bp in length
 * 9052 9151: gap of unknown length
 * 9152 10243: contig of 1092 bp in length
 * 10244 10343: gap of unknown length
 * 10344 12294: contig of 1951 bp in length
 * 12295 12394: gap of unknown length
 * 12395 13603: contig of 1209 bp in length
 * 13604 15416: contig of 1713 bp in length
 * 15417 15516: gap of unknown length

15517 17145: contig of 1629 bp in length
 * 17146 17245: gap of unknown length
 * 17246 18333: contig of 1088 bp in length
 * 18334 18433: gap of unknown length
 * 18434 20132: contig of 1659 bp in length
 * 20133 20232: gap of unknown length
 * 20233 22222: contig of 1990 bp in length
 * 22223 22323: gap of unknown length
 * 22323 22387: contig of 1065 bp in length
 * 22388 23487: gap of unknown length
 * 23488 25187: contig of 1700 bp in length
 * 25188 25287: gap of unknown length
 * 25288 26760: contig of 1473 bp in length
 * 26761 26860: gap of unknown length
 * 26861 28718: contig of 1858 bp in length
 * 28719 28818: gap of unknown length
 * 28819 30780: contig of 1962 bp in length
 * 30781 30880: gap of unknown length
 * 30881 32569: contig of 1669 bp in length
 * 32570 32659: gap of unknown length
 * 32660 35992: contig of 3323 bp in length
 * 35993 36092: gap of unknown length
 * 36093 36093: contig of 2935 bp in length
 * 36094 39047: gap of unknown length
 * 39048 39147: gap of unknown length
 * 39148 41294: contig of 2147 bp in length
 * 41295 41394: gap of unknown length
 * 41395 43616: contig of 2222 bp in length
 * 43617 43716: gap of unknown length
 * 43717 45880: contig of 2164 bp in length
 * 45881 45980: gap of unknown length
 * 45981 49866: contig of 3866 bp in length
 * 49867 49965: gap of unknown length
 * 49966 52901: contig of 2933 bp in length
 * 52902 53001: gap of unknown length
 * 53002 57026: contig of 4023 bp in length
 * 57027 57126: gap of unknown length
 * 57127 59784: contig of 2658 bp in length
 * 59785 59884: gap of unknown length
 * 59885 63872: contig of 3988 bp in length
 * 63873 63972: gap of unknown length
 * 63973 68976: contig of 5004 bp in length
 * 68977 69076: gap of unknown length
 * 69077 74024: contig of 4948 bp in length
 * 74025 74124: gap of unknown length
 * 74125 80131: contig of 6007 bp in length
 * 80132 80231: gap of unknown length
 * 80232 85840: contig of 5609 bp in length
 * 85841 85940: gap of unknown length
 * 85941 91519: contig of 5579 bp in length
 * 91520 91620: gap of unknown length
 * 91621 97625: contig of 6006 bp in length
 * 97626 97725: gap of unknown length
 * 97726 103912: contig of 6187 bp in length
 * 103913 104012: gap of unknown length
 * 104013 109839: contig of 5827 bp in length
 * 109840 109939: gap of unknown length
 * 109940 117007: contig of 7068 bp in length
 * 117008 117107: gap of unknown length
 * 117108 126809: contig of 9702 bp in length.
 * location/Qualifiers
 1. 126809
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-392N11"

BASE COUNT 34157 a 26109 c 25934 g 34674 t 5935 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1.21e+03
 Score: 8.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 53.33%
 Length: 126809
 Matches: 8
 Conservative: 0
 Mismatches: 0
 Indels: 0

DB: 2 Gaps: 0

SEQ1-65to79 (1-15) x AC124138 (1-126809)

QY 1 SerleuArgSerAlaHisLeuAla 8
 |||||||

Db 40908 AGCTCAGAGATGCCACCTGGCC 40931

RESULT 44
 AL157764 137910 bp DNA linear PRI 13-SEP-2001
 LOCUS Human DNA sequence from clone RP11-562E17 on chromosome 13,
 DEFINITION complete sequence.
 AC157764
 VERSION AL157764 GI:15620590
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 137910)

REFERENCE
 AUTHORS Smith, M.
 TITLE Direct Submission
 JOURNAL Submitted (13-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequests@sanger.ac.uk
 On Sep 14, 2001 this sequence version replaced gi:15131959.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw,
 SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 RP11-562E17 is from the library RPCI-11.2 constructed by the group
 of Pletier de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-562E17. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true right end of clone RP11-562E17 is at 137910 in this
 sequence. The true right end of clone RP11-21F17 is at 2000 in
 this sequence.

FEATURES
 Source Location/Qualifiers
 1..137910
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-562E17"
 /clone_id="RPCI-11.2"
 BASE COUNT 44439 a 23798 c 24651 g 45022 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.31e+03 Length: 137910
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 9 Gaps: 0

SEQ1-65to79 (1-15) x AL157764 (1-137910)

QY 1 SerleuArgSerAlaHisLeuAla 8
 |||||||

Db 39902 AGTTGAGAGAGTGCATCTGCCA 39925

RESULT 45
 AC108328 147108 bp DNA linear HTG 13-JUL-2002
 LOCUS Rattus norvegicus clone CH230-303C16, *** SEQUENCING IN PROGRESS
 DEFINITION ** 65 unordered pieces.
 AC108328
 VERSION AC108328.3 GI:21737630
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 147108)

REFERENCE
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbarella, J., Benton, J., Bimge, K., Blankenburg, K., Bonnin, D.,
 Bonck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buha, J., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Degen, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Huik, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louiseged, H.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhinney, E., McLeod, M.P., Meadow, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, K., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokkwo, S., Ogun, M., Okunolu, G.,
 Orangun, N., Owiedo, R., Pace, A., Payton, B., Peety, J., Peters, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Qules, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoshbar, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabori, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, C., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 147108)
 TITLE
 JOURNAL
 AUTHORS
 TITLE
 JOURNAL
 Direct Submission
 Submitted (27-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 147108)

AUTHORS
TITLE
JOURNAL
COMMENT

Morley, K.C.
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced g1:18846586.

Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPMF
Center clone name: CH230-303C16
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 85923 bases at least Q40
Consensus quality: 89551 bases at least Q30
Consensus quality: 91900 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence,
* as soon as it is available and the accession number will
* be preserved.

1 1027: contig of 1027 bp in length
* 1028 1137: gap of unknown length
* 1128 2314: contig of 1187 bp in length
* 2315 2414: gap of unknown length
* 2415 3495: contig of 1081 bp in length
* 3496 3595: gap of unknown length
* 3596 5110: contig of 1515 bp in length
* 5111 5210: gap of unknown length
* 5211 6308: contig of 1098 bp in length
* 6309 6408: gap of unknown length
* 6409 7498: contig of 1090 bp in length
* 7499 7598: gap of unknown length
* 7599 9103: contig of 1505 bp in length
* 9104 9203: gap of unknown length
* 9204 10509: contig of 1306 bp in length
* 10510 10609: gap of unknown length
* 10610 11817: contig of 1208 bp in length
* 11818 11917: gap of unknown length
* 11919 13391: contig of 1474 bp in length
* 13392 13491: gap of unknown length
* 13492 14940: contig of 1449 bp in length
* 14941 15040: gap of unknown length
* 15041 16295: contig of 1255 bp in length
* 16296 16395: gap of unknown length
* 16396 18171: contig of 1776 bp in length
* 18172 18271: gap of unknown length
* 18272 19589: contig of 1318 bp in length
* 19590 19689: gap of unknown length
* 19690 21045: contig of 1356 bp in length
* 21046 21145: gap of unknown length
* 21146 22535: contig of 1390 bp in length
* 22536 22635: gap of unknown length
* 22636 24177: contig of 1542 bp in length
* 24178 24277: gap of unknown length
* 24278 25527: contig of 1250 bp in length
* 25528 25627: gap of unknown length
* 25628 26889: contig of 1262 bp in length
* 26890 26989: gap of unknown length
* 26990 28155: contig of 1166 bp in length
* 28156 28255: gap of unknown length
* 28256 29720: contig of 1465 bp in length

29721	29820:	gap of unknown length
29821	31154:	contig of 1334 bp in length
31155	31254:	gap of unknown length
31255	32397:	contig of 1143 bp in length
32398	32497:	gap of unknown length
32498	34024:	contig of 1527 bp in length
34025	34124:	gap of unknown length
34125	35965:	contig of 1841 bp in length
35966	36065:	gap of unknown length
36066	37418:	contig of 1353 bp in length
37419	37518:	gap of unknown length
37519	38707:	contig of 1189 bp in length
38708	38807:	gap of unknown length
38808	40854:	contig of 2047 bp in length
40855	40954:	gap of unknown length
40955	43162:	contig of 2208 bp in length
43163	43262:	gap of unknown length
43263	45400:	contig of 2138 bp in length
45401	45500:	gap of unknown length
45501	46903:	contig of 1403 bp in length
46904	47003:	gap of unknown length
47004	48067:	contig of 1064 bp in length
48068	48167:	gap of unknown length
48168	50089:	contig of 1922 bp in length
50090	50189:	gap of unknown length
50190	52640:	contig of 2451 bp in length
52641	52740:	gap of unknown length
52741	54289:	contig of 1549 bp in length
54290	54389:	gap of unknown length
54390	55471:	contig of 1082 bp in length
55472	55571:	gap of unknown length
55572	57630:	contig of 2059 bp in length
57631	57730:	gap of unknown length
57731	60710:	contig of 2980 bp in length
60711	60810:	gap of unknown length
60811	62472:	contig of 1662 bp in length
62473	62572:	gap of unknown length
62573	64553:	contig of 1981 bp in length
64554	64653:	gap of unknown length
64654	67022:	contig of 2368 bp in length
67023	67122:	gap of unknown length
67123	68896:	contig of 1774 bp in length
68897	68996:	gap of unknown length
68997	70395:	contig of 1399 bp in length
70396	70495:	gap of unknown length
70496	72405:	contig of 1910 bp in length
72406	72505:	gap of unknown length
72506	75526:	contig of 3021 bp in length
75527	75626:	gap of unknown length
75627	77345:	contig of 1719 bp in length
77346	77445:	gap of unknown length
77446	79398:	contig of 1953 bp in length
79399	79498:	gap of unknown length
79499	81984:	contig of 2468 bp in length
81985	82084:	gap of unknown length
82085	83808:	contig of 1724 bp in length
83809	83908:	gap of unknown length
83909	86374:	gap of unknown length
86375	86674:	gap of unknown length
86675	90345:	contig of 3671 bp in length
90346	90445:	gap of unknown length
90446	93750:	contig of 3205 bp in length
93651	93750:	gap of unknown length
93751	96706:	contig of 2956 bp in length
96707	96806:	gap of unknown length
96807	101061:	contig of 4255 bp in length

Alignment Scores:
Pred. No.: 1.38e+03
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 53.33%
Length: 147108
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0

DB: 2 Gaps: 0

Seq1-65to79 (1-15) x AC108328 (1-147108)

QY 1 SerLeuArgSerAlaHisLeuAla 8
 |||||

Db 27642 AGCCTCAGAGTCCACCTGAGCC 27665

RESULT 46
 AC115134 148879 bp DNA linear HTG 13-JUL-2002
 AC115134
 Rattus norvegicus clone CH230-245C6, *** SEQUENCING IN PROGRESS
 *** 59 unordered pieces.

DEFINITION
 AC115134
 AC115134.3 GI:21735868
 HTG: HTG.PHASE1.
 Norway rat.
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 148879)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Alstbrooks, S.L., Amaralunga, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbarta, J., Benton, J., Binnage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowe, S., Bileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Delavey, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douhaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagay, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudh, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Lead, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtharge, O., Liew, C., Liu, J., Liu, W., Louised, H.,
 Lozdo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, E., Nwokenkwo, S., Oguh, M., Nguyen, G.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Nguyen, N.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Remy, Y.,
 Rivers, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshani, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wlecezyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE

2 (bases 1 to 148879)
 Direct Submission
 Worley, K.C.
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 148879)
 Worley, K.C.
 Direct Submission

JOURNAL COMMENT

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:21239880.

Genome Center

Center: Baylor College of Medicine
 Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

Center project name: GMD
 Center clone name: CH230-245C6

Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap: version 0.990329
 Consensus quality: 107621 bases at least Q40

Consensus quality: 111452 bases at least Q30
 Consensus quality: 113094 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently
 consists of 59 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1264: contig of 1264 bp in length
 1265 1364: gap of unknown length
 1365 2394: contig of 1030 bp in length
 2395 2494: gap of unknown length
 2495 3719: contig of 1225 bp in length
 3720 3820: gap of unknown length
 3820 4821: contig of 1002 bp in length
 4821 4921: gap of unknown length
 4921 5990: contig of 1069 bp in length
 5990 6091: gap of unknown length
 6091 7240: contig of 1150 bp in length
 7240 7341: gap of unknown length
 7341 8356: contig of 1016 bp in length
 8356 8457: gap of unknown length
 8457 9714: contig of 1258 bp in length
 9714 9814: gap of unknown length
 9814 11187: contig of 1373 bp in length
 11187 11287: gap of unknown length
 11287 12336: contig of 1049 bp in length
 12336 12437: gap of unknown length
 12437 13869: contig of 1433 bp in length
 13869 13970: gap of unknown length
 13970 15035: contig of 1066 bp in length
 15035 15135: gap of unknown length
 15135 16761: contig of 1626 bp in length
 16761 18470: gap of unknown length
 18470 18570: contig of 1609 bp in length
 18570 20002: gap of unknown length
 20002 20102: contig of 1432 bp in length
 20102 21435: gap of unknown length
 21435 21535: contig of 1333 bp in length
 21535 23238: gap of unknown length
 23238 23339: contig of 1703 bp in length
 23339 24592: gap of unknown length
 24592 24693: contig of 1254 bp in length
 24693 26402: gap of unknown length
 26402 26502: contig of 1710 bp in length
 26502 28724: gap of unknown length
 28724 28824: contig of 2222 bp in length
 28824 30766: gap of unknown length
 30766 30866: contig of 1942 bp in length
 30866 33323: contig of 2457 bp in length

*	33324	33423:	gap of unknown length
*	33424	34930:	contig of 1507 bp in length
*	34931	35030:	gap of unknown length
*	35031	36544:	contig of 1514 bp in length
*	36545	36644:	gap of unknown length
*	36645	39516:	contig of 2872 bp in length
*	39517	39616:	gap of unknown length
*	39617	41543:	contig of 1927 bp in length
*	41544	41643:	gap of unknown length
*	41644	44083:	contig of 2440 bp in length
*	44084	44183:	gap of unknown length
*	44184	46366:	contig of 2183 bp in length
*	46367	46466:	gap of unknown length
*	46467	48564:	contig of 2098 bp in length
*	48565	48664:	gap of unknown length
*	48665	50192:	contig of 1528 bp in length
*	50193	50292:	gap of unknown length
*	50293	52072:	contig of 1780 bp in length
*	52073	52172:	gap of unknown length
*	52173	54964:	contig of 2792 bp in length
*	54965	55064:	gap of unknown length
*	55065	57063:	contig of 1999 bp in length
*	57064	57163:	gap of unknown length
*	57164	59656:	contig of 2493 bp in length
*	59657	59756:	gap of unknown length
*	59757	62837:	contig of 3081 bp in length
*	62838	62937:	gap of unknown length
*	62938	65600:	contig of 2663 bp in length
*	65601	65700:	gap of unknown length
*	65701	67327:	contig of 1627 bp in length
*	67328	67427:	gap of unknown length
*	67428	69434:	contig of 2007 bp in length
*	69435	69534:	gap of unknown length
*	69535	72733:	contig of 3199 bp in length
*	72734	72833:	gap of unknown length
*	72834	75880:	contig of 2747 bp in length
*	75881	75980:	gap of unknown length
*	75981	78401:	contig of 2721 bp in length
*	78402	78501:	gap of unknown length
*	78502	81176:	contig of 2675 bp in length
*	81177	81276:	gap of unknown length
*	81277	84098:	contig of 2822 bp in length
*	84099	84198:	gap of unknown length
*	84199	88407:	contig of 4209 bp in length
*	88408	88507:	gap of unknown length
*	88508	92196:	contig of 3689 bp in length
*	92197	92296:	gap of unknown length
*	92297	94544:	contig of 2248 bp in length
*	94545	94644:	gap of unknown length
*	94645	98393:	contig of 3749 bp in length
*	98394	98493:	gap of unknown length
*	98494	102471:	contig of 3978 bp in length
*	102472	102571:	gap of unknown length
*	102572	105277:	contig of 2706 bp in length
*	105278	105377:	gap of unknown length
*	105378	109088:	contig of 3711 bp in length
*	109089	109188:	gap of unknown length
*	109189	111789:	contig of 2601 bp in length
*	111790	111889:	gap of unknown length
*	111890	115231:	contig of 3342 bp in length
*	115232	115331:	gap of unknown length
*	115332	119446:	contig of 4115 bp in length
*	119447	119546:	gap of unknown length
*	119547	123052:	contig of 3506 bp in length

Alignment Scores:
 Pred. No.:
 Score:
 Percent Similarity:
 Best Local Similarity:
 Query Match:
 DB:

1.4e+03
 8.00
 100.00%
 100.00%
 53.33%
 2

Length: 148879
 Matches: 8
 Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0

SEQ1-65T079 (1-15) x AC115134 (1-148879)
 QY 1 SerLeuArgSerAlaHisLeuAla 8
 Db 81556 AGCCTCAGAGTCCACCTGACC 81579
 RESULT 47
 AL354740/c
 LOCUS
 DEFINITION
 Human DNA sequence from clone RP11-513115 on chromosome 6, complete sequence.
 ACCESSION
 AL354740
 VERSION
 AL354740.29 GI:15020744
 KEYWORDS
 HTG.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 151828)
 REFERENCE
 AUTHORS
 Williams,S.
 TITLE
 Direct Submision
 JOURNAL
 Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jul 25, 2001 this sequence version replaced gi:14626095.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: EM: EMBL, SW:
 SWISSPROT; TR: TrEMBL; WP: WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormpep
 This sequence
 was generated from part of bacterial clone configs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr6
 RP11-513115 is from the library RPCT-11.2 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-513115 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-513115 is at 1 in this sequence.
 The true left end of clone RP1-187N21 is at 149829 in this
 sequence. The true right end of clone RP13-459B11 is at 22924 in
 this sequence.

FEATURES

source

1..151828
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-513115"
 /clone_1lb="RPCT-11.2"
 1538..1622
 /note="MIR repeat: matches 82. .165 of consensus"
 2003..2078
 /note="19 copies 4 mer cctc 71% conserved"
 2402..2438
 /note="MIR repeat: matches 95. .131 of consensus"
 2741..2833
 /note="MIR repeat: matches 49. .145 of consensus"

```
repeat_region 3441. .3599 /note="MIR repeat: matches 76. .235 of consensus"
repeat_region 3903. .3942 /note="5 copies 8 mer acacacac 100% conserved"
repeat_region 4922. .5234 /note="Aluub repeat: matches 1. .301 of consensus"
repeat_region 6809. .6864 /note="L2 repeat: matches 2438. .2493 of consensus"
repeat_region 7209. .7816 /note="L2 repeat: matches 2137. .2743 of consensus"
misc_feature 11298. .13506 /note="Cpg island"
evidence-not_experimental 12892. .12951 /note="30 copies 2 mer ag 71% conserved"
repeat_region 12901. .12948 /note="4 copies 12 mer 89% conserved"
repeat_region 12903. .12950 /note="6 copies 8 mer gagcagac 81% conserved"
repeat_region 13169. .13316 /note="74 copies 2 mer cc 59% conserved"
repeat_region 13560. .13621 /note="MIR repeat: matches 80. .141 of consensus"
repeat_region 14337. .14423 /note="L2 repeat: matches 2626. .2710 of consensus"
repeat_region 14797. .14914 /note="MIR repeat: matches 34. .169 of consensus"
repeat_region 14915. .15111 /note="MER9A repeat: matches 1. .195 of consensus"
repeat_region 15240. .15279 /note="5 copies 8 mer tgaatgaa 92% conserved"
repeat_region 15688. .15795 /note="9 copies 12 mer 79% conserved"
repeat_region 15689. .15792 /note="13 copies 8 mer acacacac 80% conserved"
repeat_region 17157. .17454 /note="MER46C repeat: matches 12. .338 of consensus"
repeat_region 17921. .18050 /note="2 copies 65 mer 97% conserved"
repeat_region 20310. .20391 /note="MIR repeat: matches 108. .185 of consensus"
repeat_region 20943. .21281 /note="Alusx repeat: matches 1. .299 of consensus"
repeat_region 21489. .21562 /note="MIR repeat: matches 102. .180 of consensus"
repeat_region 21871. .21922 /note="13 copies 4 mer tggg 73% conserved"
repeat_region 21930. .21969 /note="5 copies 8 mer acacacac 97% conserved"
misc_feature 22038. .22511 /note="Cpg island"
evidence-not_experimental 22774. .22866 /note="MER5B repeat: matches 1. .101 of consensus"
repeat_region 22867. .22986 /note="FLAM_C repeat: matches 1. .129 of consensus"
repeat_region 22987. .23059 /note="MER5B repeat: matches 101. .171 of consensus"
repeat_region 23436. .23731 /note="Alusx repeat: matches 1. .311 of consensus"
repeat_region 23732. .23759 /note="7 copies 4 mer gaaa 92% conserved"
repeat_region 24098. .24357 /note="LIME3 repeat: matches 5833. .6088 of consensus"
repeat_region 24392. .24697 /note="Aluuo repeat: matches 5. .302 of consensus"
repeat_region 25606. .25889 /note="Aluub repeat: matches 1. .269 of consensus"
repeat_region 26309. .26504 /note="Aluub repeat: matches 90. .302 of consensus"
repeat_region 26575. .26890 /note="Alusx repeat: matches 1. .312 of consensus"
repeat_region 27027. .27319 /note="Aluuo repeat: matches 3. .296 of consensus"
repeat_region 27586. .27736 /note="LIMC4 repeat: matches 7237. .7390 of consensus"
repeat_region 27737. .28033 /note="AluY repeat: matches 1. .298 of consensus"
repeat_region 28034. .28085 /note="LIMC4 repeat: matches 7389. .7440 of consensus"
repeat_region 28086. .28205 /note="10 copies 12 mer 80% conserved"
repeat_region 28208. .28325 /note="FLAM_A repeat: matches 1. .119 of consensus"
repeat_region 28333. .28344 /note="Alusg/x repeat: matches 121. .132 of consensus"
repeat_region 28345. .28637 /note="Alusx repeat: matches 1. .297 of consensus"
repeat_region 28638. .28816 /note="Alusg/x repeat: matches 131. .312 of consensus"
repeat_region 29512. .29659 /note="L2 repeat: matches 1995. .2150 of consensus"
repeat_region 30114. .30291 /note="Alusx repeat: matches 151. .312 of consensus"
repeat_region 30292. .30593 /note="Alusg1 repeat: matches 1. .302 of consensus"
repeat_region 30594. .30739 /note="Alusx repeat: matches 1. .151 of consensus"
repeat_region 31398. .31713 /note="Alusg repeat: matches 3. .308 of consensus"
repeat_region 31740. .31938 /note="LIM2 repeat: matches 6110. .6308 of consensus"
repeat_region 31939. .32234 /note="Alusx repeat: matches 3. .304 of consensus"
repeat_region 32235. .34816 /note="LIM2 repeat: matches 3482. .6110 of consensus"
repeat_region 34817. .34979 /note="MSTA repeat: matches 1. .168 of consensus"
repeat_region 34980. .35274 /note="Alusp repeat: matches 1. .296 of consensus"
repeat_region 35275. .35552 /note="MSTB repeat: matches 168. .426 of consensus"
repeat_region 35553. .36313 /note="LIM2 repeat: matches 2720. .3482 of consensus"
repeat_region 36315. .36618 /note="Alusx repeat: matches 3. .306 of consensus"
repeat_region 36621. .36787 /note="Alusx repeat: matches 134. .297 of consensus"
repeat_region 36788. .37076 /note="Alusx repeat: matches 24. .312 of consensus"
repeat_region 37077. .37209 /note="Alusx repeat: matches 1. .134 of consensus"
repeat_region 37221. .37795 /note="LIM1 repeat: matches 1824. .2773 of consensus"
repeat_region 38091. .38138 /note="4 copies 12 mer 87% conserved"
repeat_region 38094. .38143 /note="25 copies 2 mer aa 72% conserved"
repeat_region 38190. .38503 /note="Alusx repeat: matches 1. .312 of consensus"
repeat_region 38504. .38613 /note="55 copies 2 mer ga 65% conserved"
repeat_region 38523. .38614 /note="23 copies 4 mer agga 81% conserved"
repeat_region 38619. .39430

Alignment Scores:
Pred. No.: 1.42e+03
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 53.33%
DB: 9
Gaps: 0

SEQ1-65to79 (1-15) x AL354740 (1-151828)
```

Oy 4 SerIahstleuAlaglyCInser 11
 Db 14747 AGTCACACTTGGCTGGCAGAGC 14724
 RESULT 48
 AC119474
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-424C21, *** SEQUENCING IN PROGRESS
 AC119474 152191 bp DNA linear HTG 18-JUL-2002
 AC119474
 VERSION AC119474.5 GI:21746905
 KEYWORDS HTG, HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 152191)
 REFERENCE
 AUTHORS
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbieri,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Davis,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dlin,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,V., Hulik,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
 Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Meshawari,M., Mapua,P., Martin,R., Matlindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okunom,G.,
 Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherger,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,I.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,K., Tang,H.,
 Tansy,J., Taylor,C., Taylor,T., Telford,B., Tamerisa,K., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 TITLE
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 152191)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 3 (bases 1 to 152191)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Jul 14, 2002 this sequence version replaced gi:20799930.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GVPD
 Center clone name: CH230-424C21
 ----- Summary Statistics
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.950329
 Consensus quality: 124262 bases at least Q40
 Consensus quality: 128161 bases at least Q30
 Consensus quality: 131092 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 34 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1
 1325 1324: contig of 1324 bp in length
 1424: gap of unknown length
 1425 2449: contig of 1025 bp in length
 2450 2549: gap of unknown length
 2550 3884: contig of 1335 bp in length
 3885 3984: gap of unknown length
 3985 5092: contig of 1108 bp in length
 5093 5192: gap of unknown length
 5193 6317: contig of 1125 bp in length
 6318 6417: gap of unknown length
 6418 7979: contig of 1562 bp in length
 7980 8079: gap of unknown length
 8080 9512: contig of 1433 bp in length
 9513 9612: gap of unknown length
 9613 10830: contig of 1218 bp in length
 10831 10930: gap of unknown length
 10931 11979: contig of 1049 bp in length
 11980 12079: gap of unknown length
 12080 14117: contig of 2038 bp in length
 14118 14217: gap of unknown length
 14218 16380: contig of 2173 bp in length
 16381 16490: gap of unknown length
 16491 17821: contig of 1331 bp in length
 17822 17921: gap of unknown length
 17922 20794: contig of 2873 bp in length
 20795 20894: gap of unknown length
 20895 24072: gap of 3178 bp in length
 24073 24172: gap of unknown length
 24173 27441: contig of 3269 bp in length
 27442 27541: gap of unknown length
 27542 30052: contig of 2511 bp in length
 30053 30152: gap of unknown length
 30153 32002: contig of 1850 bp in length
 32003 32102: gap of unknown length
 32103 33684: contig of 1582 bp in length
 33685 33784: gap of unknown length
 33785 37287: contig of 3503 bp in length
 37288 37387: gap of unknown length
 37388 41765: gap of 4378 bp in length
 41766 41865: gap of unknown length
 41866 45232: contig of 3367 bp in length
 45233 45332: gap of unknown length
 45333 48372: contig of 3040 bp in length
 48373 48472: gap of unknown length
 48473 53578: contig of 5106 bp in length
 53579 53678: gap of unknown length
 53679 58379: contig of 4658 bp in length

* 58337 58436: gap of unknown length
 * 58437 64553: contig of 6117 bp in length
 * 64554 64554: gap of unknown length
 * 64554 69547: contig of 4894 bp in length
 * 69548 69647: gap of unknown length
 * 69648 74136: contig of 4489 bp in length
 * 74137 74236: gap of unknown length
 * 74237 80961: contig of 6725 bp in length
 * 80962 81061: gap of unknown length
 * 81062 88304: contig of 7243 bp in length
 * 88305 88404: gap of unknown length
 * 88405 95126: contig of 6722 bp in length
 * 95127 95226: gap of unknown length
 * 95227 103086: contig of 7860 bp in length
 * 103087 103186: gap of unknown length
 * 103187 112587: contig of 9401 bp in length
 * 112588 112687: gap of unknown length
 * 112688 128551: contig of 15864 bp in length
 * 128552 128651: gap of unknown length
 * 128652 152191: contig of 23540 bp in length.

FEATURES
 source
 1.152191
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-424C21"

BASE COUNT 42875 a 27225 c 26942 g 49377 t 5772 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1.42e+03 Length: 152191
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 2 Gaps: 0

SEQ1-65to79 (1-15) x AC119474 (1-152191)

QY 1 SerleuArgerSerIahSleuAa 8
 Db 108929 AGCCTCAGAGATCCCACTGGCC 108952

RESULT 49
 AL355871/c
 LOCUS AL355871 154588 bp DNA linear PRI 01-AUG-2000
 DEFINITION Human DNA sequence from clone RP11-47K11 on chromosome 1, complete
 sequence.
 ACCESSION AL355871
 VERSION AL355871.5 GI:9407857
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 154588)
 AUTHORS Corby,N.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 REQUESTERS: clonerequest@sanger.ac.uk
 On Jul 23, 2000 this sequence version replaced gi:9368089.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 EMBL, SW, SWISSPROT, Tr, TREMBL, WP, WORMPEP, Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/MGP/Chr1>
 RP11-47K11 is from the library RP11-11 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pBACe3.6
 This sequence is the entire insert of clone RP11-47K11.

FEATURES
 source
 1.154588
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-47K11"
 /clone_1bp="RP11-11"
 71035
 /note="Tandem repeat. Forced join. Gap size estimated to
 be approximately 1.1kb by BAMHI, ECORI and HINDIII
 restriction enzyme digest data."

BASE COUNT 46359 a 31281 c 29218 g 47730 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.44e+03 Length: 154588
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 9 Gaps: 0

SEQ1-65to79 (1-15) x AL355871 (1-154588)

QY 8 AlaglyGInserIleuSergly 15
 Db 69571 GCAGGACAAAGCATCTATCTGGG 69548

RESULT 50
 AC004846/c
 LOCUS AC004846 154616 bp DNA linear PRI 08-NOV-2000
 DEFINITION Homo sapiens PAC clone RP4-647C14 from 14q24.3, complete sequence.
 ACCESSION AC004846
 VERSION AC004846.2 GI:7243869
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 154616)
 AUTHORS Sulston,J.E. and Waterston,R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 154616)
 AUTHORS Ryan,E., Sun,H. and Spalding,L.
 TITLE The sequence of Homo sapiens PAC clone RP4-647C14
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 154616)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 154616)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (15-MAR-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 154616)
AUTHORS
TITLE
JOURNAL
Submitted (14-OCT-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 154616)
AUTHORS
TITLE
JOURNAL
Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 15, 2000 this sequence version replaced g1:3213151.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplenswatson.wustl.edu
----- Summary Statistics
Center project name: H_DJ0647C14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone from chromosome 14 was provided by Dr. Pieter de Jong, Roswell Park Cancer Institute, Human Genetics Department, Elm and Carlton Streets, Buffalo NY 14263-0001 USA.

SOURCE INFORMATION:
This clone was derived from human PAC library RPCT-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP1-240K6; the clone sequenced to the right is RP4-687K1. Actual start of this clone is at base position 197 of RP4-647C14; actual end is at base position 730 of RP4-687K1.

The sequence RP4-647C14 contains a dinucleotide AT repeat from base position 4427 to 4478. The exact number of copies could not be determined.

The sequence from position 141555 to 141760 was derived from PCR product of RP4-647C14 PAC DNA.
location/Qualifiers
1..154616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3"
/clone="RP4-647C14"
/clone_1lb="RPCT-4"
1..251
repeat_region
/rpt_family="Alu"
repeat_region
288..319

/rpt_family="MER113"
320..595
repeat_region
/rpt_family="Alu"
596..627
repeat_region
/rpt_family="MER113"
1187..1330
misc_feature
/note="similar to Mus musculus EST A1414568 (NID:g4258072)
ma48903.x1"
1187..1330
misc_feature
/note="similar to Mus musculus EST A1605646 (NID:g4614813)
ma48903.y1"
2118..2410
repeat_region
/rpt_family="Alu"
2425..2728
repeat_region
/rpt_family="Alu"
2993..3023
repeat_region
/rpt_family="AT-rich"
3222..3248
repeat_region
/rpt_family="(TTTA)n"
3249..3390
repeat_region
/rpt_family="Alu"
3498..3569
repeat_region
/rpt_family="L1"
3570..3883
repeat_region
/rpt_family="Alu"
3884..3908
repeat_region
/rpt_family="L1"
3955..4079
repeat_region
/rpt_family="L2"
4099..4114
repeat_region
/rpt_family="MER1-type"
4115..4425
repeat_region
/rpt_family="Alu"
4426..4491
repeat_region
/rpt_family="MER1-type"
4492..4522
repeat_region
/rpt_family="(TC)n"
4546..4711
repeat_region
/rpt_family="L2"
4807..5214
misc_feature
/note="match to EST H73219 (NID:g1046806) ys12c01.r1"
4889..4937
repeat_region
/rpt_family="L1"
4938..4978
repeat_region
/rpt_family="(TTTTA)n"
4979..5260
repeat_region
/rpt_family="Alu"
5261..5368
repeat_region
/rpt_family="L1"
5436..5456
repeat_region
/rpt_family="AT-rich"
5457..5763
repeat_region
/rpt_family="Alu"
5764..5784
repeat_region
/rpt_family="AT-rich"
5792..5937
repeat_region
/rpt_family="L1"
5948..6127
repeat_region
/rpt_family="L2"
6128..6409
repeat_region
/rpt_family="Alu"
6410..6480
repeat_region
/rpt_family="L2"
6484..6782
repeat_region
/rpt_family="Alu"
6853..6912
repeat_region
/rpt_family="MER4-group"
6913..7050
repeat_region
/rpt_family="Alu"
7051..7357
repeat_region
/rpt_family="Alu"
7363..7649
repeat_region
/rpt_family="Alu"

```

misc_feature      7363..7378      /note="match to EST AA760631 (NID:g2809561) nz16b05.s1"
misc_feature      7363..7378      /note="match to EST AI373438 (NID:g4153304) qz94b08.x1"
misc_feature      7363..7375      /note="match to EST AA534704 (NID:g2278957) nf75d10.s1"
repeat_region     7650..7795      /rpt_family="Alu"
repeat_region     8221..8560      /rpt_family="MER1_type"
repeat_region     8981..9083      /rpt_family="MER2_type"
repeat_region     9092..9459      /rpt_family="L2"
repeat_region     9496..9671      /rpt_family="L1"
repeat_region     9771..9820      /rpt_family="L1"
repeat_region     9825..9956      /rpt_family="Alu"
repeat_region     10049..10282      /rpt_family="L2"
repeat_region     10458..10757      /rpt_family="Alu"
repeat_region     10899..11021

```

Alignment Scores:

```

Pred. No.:      1.44e+03      Length:      154616
Score:          8.00          Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     53.33%          Indels:      0
DB:              9            Gaps:          0

```

SEQ1-65to79 (1-15) x AC004846 (1-154616)

```

Qy      3 ArgSerAlaHisLeuAlaGlyIn 10
        |||||||||||||||||||
Db 86770 CGCTCGCCCACTGTGGCGGCAG 86747

```

Search completed: January 28, 2003, 12:15:22
Job time : 1666.67 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:32:26 ; Search time 344.333 Seconds
(without alignments)
98.103 Million cell updates/sec

Title: SEQ1-65TO79
Perfect score: 15
Sequence: 1 SURSAHLAQSTILSC 15

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 8

Total number of hits satisfying chosen parameters: 28

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODE=frame+ p2n model -DEV=ylh
-Q/cgnt2.1/USPTO.spool/BORIN682/runat_23012003.130140.7866/app.query.fasta_1.597
-DB=N_geneseq_101002 -OPMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPEXT=0
-DOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=8 -ALIGN=50 -MODE=LOCAL
-OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=BORIN682.ecgn.1.1.187_@runat_23012003.130140.7866 -NCPU=6 -TCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEVTIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YCAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
1: N_geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	723	24	ABK50864	Escherichia coli h
2	15	100.0	777	6	AAAS0205	Sequence of the p1
3	15	100.0	777	6	AAAS0206	Sequence of the p1
4	15	100.0	782	21	AAAS1106	Plant-optimized E.
5	15	100.0	782	21	AAAS1147	Plant-optimized E.
6	15	100.0	782	21	AAAS1944	Plant-optimized E.
7	15	100.0	1143	24	ABL40640	E. coli mutant hea
8	15	100.0	1148	13	AAO23864	Thermolabile toxin
9	15	100.0	1152	24	ABL40639	E. coli heat-labli
10	15	100.0	1508	22	AAAS01506	E. coli heat-labli
11	15	100.0	1514	22	AAAS01505	E. coli heat-labli
12	12	80.0	40	21	AAAS1140	Oligomer SLT-A Rev
13	10	66.7	711	14	AAO51314	Encodes Asp-53 E.C
14	10	66.7	711	14	AAO51315	Encodes Glu-53 E.C
15	10	66.7	711	14	AAO51316	Encodes Tyr-53 E.C
16	10	66.7	711	14	AAO51317	Encodes Lys-63 E.C
17	10	66.7	711	14	AAO51318	Encodes Lys-97 E.C
18	10	66.7	711	14	AAO51319	Encodes Tyr-97 E.C
19	10	66.7	711	14	AAO51320	Encodes Lys-107 E.C
20	10	66.7	711	14	AAO51321	Encodes Asp-104 E.C
21	10	66.7	711	14	AAO51322	Encodes Ser-104 E.C
22	10	66.7	711	14	AAO51323	Encodes Ser-106 E.C
23	10	66.7	711	14	AAO51324	Encodes Lys-114 E.C
24	10	66.7	711	14	AAO51325	Encodes Lys-114 E.C
25	10	66.7	711	14	AAO51326	E. coli heat labli
26	10	66.7	711	14	AAO42768	E. coli heat labli
27	10	66.7	711	20	AAV81595	Oligomer SLT-A For
28	8	53.3	40	21	AAAS1114	

ALIGNMENTS

RESULT 1
ABK50864
ID ABK50864 standard; DNN: 723 BP.
AC
XX
AC ABK50864;
XX
DT 30-JUL-2002 (first entry)
XX
XX
DE Escherichia coli heat labile enterotoxin sub-unit A.
XX
XX
KW Co-expression DNA vaccine: antibacterial; antiviral; antiparasitic;
KW immunostimulant; vaccine; immune response; systemic tolerance;
KW Tat-mediated immune deviation; ds; heat labile enterotoxin sub-unit A.
OS Escherichia coli.
XX
XX
PN WO200219968-A2.
XX
PD 14-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US28365.
XX
XX
PR 08-SEP-2000; 2000US-231070P.
XX
PR 08-SEP-2000; 2000US-231376P.
XX
PR 08-SEP-2000; 2000US-231403P.
XX
XX
XX 08-SEP-2000; 2000US-231449P.
XX
XX
XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.
XX
XX Hone D, Lewis G, Fouts T, Bagley K, Boyson M, Obrecht C;
XX PI Shata M, Agwale S;
XX
XX WPI; 2002-383031/41.
XX
XX
XX Co-expression DNA vaccines comprising an antigen-encoding region and a

PT biologically active component-encoding region, useful as vaccines
PT against viral, bacterial and parasitic pathogens, or for enhancing
PS immune responses -
XX Example 12; Page 89; 108pp; English.
XX
CC The invention describes a new DNA vaccine comprising a region encoding
CC an antigen component and a region encoding at least one biologically
CC active component such as adjuvants, immunoregulatory peptides and
CC proteins, antisense RNAs, and catalytic RNAs. The co-expression DNA
CC vaccines are useful for vaccinating animals against viral, bacterial
CC and parasitic pathogens, for enhancing immune responses, for inducing
CC systemic tolerance, and for treating and/or preventing TAT-mediated
CC immune deviation. The co-expression DNA vaccines are capable of
CC inducing significantly stronger immune responses against vaccine
CC antigens than conventional DNA vaccines, and are also capable of
CC inducing systemic tolerance. This sequence represents the Escherichia
CC coli heat labile enterotoxin sub-unit A that up-regulates cAMP levels.
CC The enterotoxin is an immunoregulatory molecule useful in the
CC co-expression DNA vaccines described in the invention.
XX
SQ Sequence 723 BP; 238 A; 128 C; 160 G; 197 T; 0 other;
XX
Alignment Scores:
Pred. No.: 9.13e-07 Length: 723
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0
SEQ1-65to79 (1-15) x AAK50864 (1-723)
OY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleuSergly 15
Db 193 AGTTTGAGAAAGTCTCATTACGACAGCTATATATATCAGCA 237
RESULT 2
AAN50205
ID AAN50205 standard; DNA; 777 BP.
XX
AC AAN50205;
XX
DT 30-OCT-1991 (first entry)
XX
DE Sequence of the pig scours heat labile toxin (LT) LTA gene.
XX
KM Pig scours vaccine; toxin; diarrhoea; ss.
XX
OS E.coli NCIB 11932.
XX
FH Key Location/Qualifiers
FT CDS 1..777
FT /*tag= a
XX
PN EPI45486-A.
XX
PD 19-JUN-1985.
XX
PF 12-DEC-1984; 84EP-0308620.
XX
PR 12-DEC-1983; 83GB-0033131.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Hayes MV, Harford S, Ross GW;
XX
DR WPI; 1985-148358/25.
DR P-PSDB; AAP50190.
XX
PT New toxoid as inactivated form of toxin for use in vaccines - is
PT obtd. from organism transformed by gene
XX

PS Disclosure; Fig 1; 61pp; English.
XX
CC AAN50205 is the gene sequence of the natural LTA gene. The LTA gene of
CC the site directed mutant SDM1 (see AAN50206) is inactive. The
CC inventors claim a vaccine prepn. active against pig scours which
CC contains an inactivated LTA component, together with additional K88
CC antigens opt. with whole cells comprising the antigens or contg. the
CC inactivated LTA.
XX
SQ Sequence 777 BP; 255 A; 136 C; 164 G; 222 T; 0 other;
XX
Alignment Scores:
Pred. No.: 9.75e-07 Length: 777
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
SEQ1-65to79 (1-15) x AAN50205 (1-777)
OY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleuSergly 15
Db 247 AGTTTGAGAAAGTCTCATTACGACAGCTATATATCAGCA 291
RESULT 3
AAN50206
ID AAN50206 standard; DNA; 777 BP.
XX
AC AAN50206;
XX
DT 30-OCT-1991 (first entry)
XX
DE Sequence of the pig scours heat labile toxin (LT) LTA gene of the
DE site directed mutant SDM1.
XX
KM Pig scours vaccine; toxin; diarrhoea; ss.
XX
OS E.coli NCIB 11932.
XX
FH Key Location/Qualifiers
FT CDS 1..777
FT /*tag= a
FT misc_difference 235..237
FT /*tag= b
FT /*note= "TCC in native SQ"
XX
PN EPI45486-A.
XX
PD 19-JUN-1985.
XX
PF 12-DEC-1984; 84EP-0308620.
XX
PR 12-DEC-1983; 83GB-0033131.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Hayes MV, Harford S, Ross GW;
XX
DR WPI; 1985-148358/25.
DR P-PSDB; AAP50191.
XX
PT New toxoid as inactivated form of toxin for use in vaccines - is
PT obtd. from organism transformed by gene
XX
PS Example; Fig 2; 61pp; English.
XX
CC AAN50205 is the gene sequence of the natural LTA gene. The LTA gene of
CC the site directed mutant SDM1 (see AAN50206) is inactive. The
CC inventors claim a vaccine prepn. active against pig scours which
CC contains an inactivated LTA component, together with additional K88
CC antigens opt. with whole cells comprising the antigens or contg. the
CC inactivated LTA.


```

XX      SQ      Sequence 777 BP: 255 A; 135 C; 164 G; 223 T; 0 other;
Alignment Scores:
Pred. No.:      9.75e-07      Length:      777
Score:          15.00      Matches:      15
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             6      Gaps:      0

SEQ1-65TO79 (1-15) x AAN50206 (1-777)
Qy      1 SerleuArgSerAlaHisLeuAlaGlyInSerIleuSergly 15
      |||||
Db      247 AGTTGAGAGTGTCTCACTTAGCAGACAGCTCTATTATCAGCA 291

RESULT 4
AAAS1106
ID      AAAS1106 standard; cDNA; 782 BP.
XX
AC      AAAS1106;
XX
DT      26-SEP-2000 (first entry)
XX
DE      Plant-optimized E. coli heat labile toxin A subunit coding sequence.
XX
KW      Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
KM      adjuvant; anti-bacterial; ss.
XX
OS      Escherichia coli.
XX
OS      Synthetic.
XX
FH      Key      Location/Qualifiers
FT      CDS      3..782
FT      FT      /*lag= a
FT      sig_peptide 3..56
FT      FT      /product= heat-labile-toxin-A
FT      mat_peptide 57..779
FT      FT      /*lag= b
FT      FT      /*lag= c
XX
XX      WO200037609-A2.
XX
XX      29-JUN-2000.
XX
XX      PD      22-DEC-1999; 99WO-US30747.
XX
XX      PF      22-DEC-1999; 98US-0113507.
XX
XX      PR      22-DEC-1998; 98US-0113507.
XX
XX      PA      (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX      PA      (MASO/) MASON H S.
XX      PA      (ARNT/) ARNTZEN C J.
XX
XX      PI      Mason HS, Arntzen CJ;
XX
XX      DR      WPI: 2000-442653/38.
XX      DR      P-PSDB: AAY96646.
XX
XX      PT      New polynucleotides encoding LT-A or CT-A polypeptides for the
XX      PT      transformation of plant cells, useful in immunogenic compositions to
XX      PT      elicit immune responses in animals
XX
XX      PS      Example 1; Fig 1; 103pp; English.
XX
XX      This plant-codon optimized cDNA encodes a synthetic Escherichia coli
XX      heat-labile toxin (LT) A subunit (LT-A). The sequence contains
XX      plant-preferred codons and eliminates sequence motifs associated with
XX      spurious mRNA processing. A single codon insertion (GTG encoding valine)
XX      was made to accommodate the creation of a NcoI restriction site around the
XX      initiator methionine codon. Novel polynucleotides encode a mutant LT-A
XX      polypeptide or a mutant Vibrio cholerae cholera toxin (CT) A subunit
XX      (CT-A) polypeptide, which have reduced enzyme activity as compared to the

```

```

CC      wild-type LT-A or CT-A polypeptide and where at least one of the codons
CC      is altered to a plant preferred codon. The polynucleotide further
CC      comprises a nucleic acid sequence encoding LT B subunit (LT-B) or a CT B
CC      subunit (CT-B). The polynucleotides are useful for the transformation of
CC      plant cells for the production of transgenic plants to produce edible
CC      vaccines, especially oral vaccines in transgenic plants for the
CC      prophylactic or therapeutic treatment against E. coli or V. cholerae. The
CC      mutant polypeptides are also useful as adjuvants.
XX
XX      SQ      Sequence 782 BP: 225 A; 186 C; 191 G; 180 T; 0 other;
Alignment Scores:
Pred. No.:      9.81e-07      Length:      782
Score:          15.00      Matches:      15
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             21      Gaps:      0

SEQ1-65TO79 (1-15) x AAAS1106 (1-782)
Qy      1 SerleuArgSerAlaHisLeuAlaGlyInSerIleuSergly 15
      |||||
Db      252 AGCTTGAGTGTCTGCTTGCAGCAGCAATCCATCCTCAGCA 296

RESULT 5
AAAS1147
ID      AAAS1147 standard; cDNA; 782 BP.
XX
XX      AC      AAAS1147;
XX
XX      DT      26-SEP-2000 (first entry)
XX
XX      DE      Plant-optimized E. coli LT-A K63 mutant coding sequence.
XX
XX      KW      Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
XX      KM      adjuvant; anti-bacterial; S63K; ss.
XX
XX      OS      Escherichia coli.
XX
XX      OS      Synthetic.
XX
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      3..782
XX      FT      FT      /*lag= a
XX      FT      sig_peptide 3..59
XX      FT      FT      /product= heat-labile-toxin-A
XX      FT      mat_peptide 60..779
XX      FT      FT      /*lag= b
XX      FT      FT      /*lag= c
XX
XX      WO200037609-A2.
XX
XX      XX      29-JUN-2000.
XX
XX      PD      22-DEC-1999; 99WO-US30747.
XX
XX      PF      22-DEC-1999; 98US-0113507.
XX
XX      PR      22-DEC-1998; 98US-0113507.
XX
XX      PA      (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX      PA      (MASO/) MASON H S.
XX      PA      (ARNT/) ARNTZEN C J.
XX
XX      PI      Mason HS, Arntzen CJ;
XX
XX      DR      WPI: 2000-442653/38.
XX      DR      P-PSDB: AAY96647.
XX
XX      PT      New polynucleotides encoding LT-A or CT-A polypeptides for the
XX      PT      transformation of plant cells, useful in immunogenic compositions to
XX      PT      elicit immune responses in animals
XX
XX      PS      Example 2; Page -: 103pp; English.
XX

```

CC This cDNA encodes a mutant S63K Escherichia coli heat-labile toxin (LT)
 CC A subunit (LT-A). The codon at nucleotide positions 246-248 was changed
 CC from the wild-type TCC to AAG which causes a change of serine to lysine
 CC in the mature protein at residue 63. The sequence contains
 CC plant-preferred codons and eliminates sequence motifs associated with
 CC spurious mRNA processing. A single codon insertion (GTG encoding valine)
 CC was made to accommodate the creation of a NcoI restriction site around
 CC the initiator methionine codon. Novel polynucleotides encode a mutant
 CC LT-A polypeptide or a mutant Vibrio cholerae cholera toxin (CT) A
 CC subunit (CT-A) polypeptide, which have reduced enzyme activity as
 CC compared to the wild-type LT-A or CT-A polypeptide and where at least one
 CC of the codons is altered to a plant preferred codon. The polynucleotide
 CC further comprises a nucleic acid sequence encoding LT B subunit (LT-B) or
 CC a CT B subunit (CT-B). The polynucleotides are useful for the
 CC transformation of plant cells for the production of transgenic plants to
 CC produce edible vaccines, especially oral vaccines in transgenic plants
 CC for the prophylactic or therapeutic treatment against E. coli or V.
 CC cholerae. The mutant polypeptides are also useful as adjuvants.
 CC Note, this sequence does not appear in the specification. It was
 CC constructed from the wild type LT-A cDNA shown in AAA51106 which is
 CC given in Figure 1 of the specification.

SO Sequence 782 BP; 227 A; 184 C; 192 G; 179 T; 0 other;

Alignment Scores:
 Pred. No.: 9.81e-07 Length: 782
 Score: 15.00 Matches: 15
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

SEQ1-65T079 (1-15) x AAA51147 (1-782)

OY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
 DB 252 AGCTTAGGTCGTGCTCATTGGCAGACAATCCATCCTCTCAGGA 296

RESULT 6
 ID AAA51544 standard; cDNA; 782 BP.
 AC AAA51544;

DT 26-SEP-2000 (first entry)

DE Plant-optimized E. coli LT-A G192 mutant coding sequence.

KW Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
 KW adjuvant; anti-bacterial; R1926; ss.

XX Escherichia coli.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 3..782

FT sig_peptide 3..59 /product= heat-labile_toxin-A

FT mat_peptide 60..779 /tag= b

FT /tag= c

PN WO200037609-A2.

PD 29-JUN-2000.

PF 22-DEC-1999; 99WO-US30747.

PR 22-DEC-1998; 98US-0113507.

PA (BOYCE-) BOYCE THOMPSON INST PLANT RES.
 PA (MASO/) MASON H S.

PA (ARNT/) ARNTZEN C J.
 XX Mason HS, Arntzen CJ;
 XX WPI: 2000-442653/38.
 DR P-FSDB; AAY96648.

PT New polynucleotides encoding LT-A or CT-A polypeptides for the
 PT transformation of plant cells, useful in immunogenic compositions to
 PT elicit immune responses in animals

PS Example 3; Page -: 103pp; English.

CC This cDNA encodes a mutant R192G Escherichia coli heat-labile toxin (LT)
 CC A subunit (LT-A). The nucleotide at position 631 was changed from the
 CC wild-type A to G which causes a change of arginine to glycine in the
 CC mature protein at residue 192. The sequence contains plant-preferred
 CC codons and eliminates sequence motifs associated with spurious mRNA
 CC processing. A single codon insertion (GTG encoding valine) was made to
 CC accommodate the creation of a NcoI restriction site around the
 CC initiator methionine codon. Novel polynucleotides encode a mutant LT-A
 CC polypeptide or a mutant Vibrio cholerae cholera toxin (CT) A subunit
 CC (CT-A) polypeptide, which have reduced enzyme activity as compared to
 CC the wild-type LT-A or CT-A polypeptide and where at least one of the
 CC codons is altered to a plant preferred codon. The polynucleotide further
 CC comprises a nucleic acid sequence encoding LT B subunit (LT-B) or a CT B
 CC subunit (CT-B). The polynucleotides are useful for the transformation of
 CC plant cells for the production of transgenic plants to produce edible
 CC vaccines, especially oral vaccines in transgenic plants for the
 CC prophylactic or therapeutic treatment against E. coli or V. cholerae.
 CC Note, this sequence does not appear in the specification. It was
 CC constructed from the wild type LT-A cDNA shown in AAA51106 which is
 CC given in Figure 1 of the specification.

SO Sequence 782 BP; 224 A; 186 C; 192 G; 180 T; 0 other;

Alignment Scores:
 Pred. No.: 9.81e-07 Length: 782
 Score: 15.00 Matches: 15
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

SEQ1-65T079 (1-15) x AAA51544 (1-782)

OY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
 DB 252 AGCTTAGGTCGTGCTCATTGGCAGACAATCCATCCTCTCAGGA 296

RESULT 7
 ID ABL40640 standard; DNA; 1143 BP.
 AC ABL40640;

DT 17-JUN-2002 (first entry)

DE E. coli mutant heat-labile toxin (mLT) 5B-SD-1A protein encoding DNA.

KW LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant;
 KW vaccine; mutant; mLT; ds.

XX Escherichia coli.

XX Key Location/Qualifiers

FT CDS 1..1143

FT /tag= a /note= "contains an internal SD sequence"

FT misc_feature 1..312 /tag= b
 FT /note= "mLT 5B subunit"

seq1-65to79.olip2n.rng

07-AUG-2000: 2000JP-0238740.
07-AUG-2000: 2000JP-0238740.
DB: 622 AGTTTGGAAGTCTCATTACGACGACAGTATTATTATCATCGA 666
SEQ1-65T079 (1-15) x ABLA0640 (1-1143)
QY 1 SerLeuArgSerAlaHisLeuAlaGlyInSerIleLeuSergly 15
DB 622 AGTTTGGAAGTCTCATTACGACGACAGTATTATTATCATCGA 666
RESULT 8
AAQ23864
ID AAQ23864 standard; DNA: 1148 BP.
XX AAQ23864:
XX
XX 08-OCT-1992 (first entry)
XX
XX Thermolabile toxin (Lth) gene.
XX
XX Lth: ST1a; ST1b; thermolabile; heat-stable; E.coli: T4DNA ligase;
XX KW probe; detection; ss.
XX
XX Escherichia coli.
XX OS
XX JP04079898-A.
XX
XX 13-MAR-1992.
XX
XX

```
DE      23-JUL-1990;          90JP-0194208.
PR      23-JUL-1990;       90JP-0194208.
XX
PA      (KITA) KITASATO RES INST.
XX
DR      WPI, 1992-137930/17.
XX
PT      DNA and RNA probe for simultaneous detection of Vibrio cholerae
XX      and enterotoxigenic E. coli
XX
XX      @chain 13 page 1; 15pp; Japanese.
XX
CC      The thermolabile toxin (LTn) gene sequence is derived from
CG      enterotoxigenic E. coli. An LTn-STn-SBP probe, wherein all
CC      all or part of the sequences of the three toxin genes represented
CC      in AAQ23864-66 are ligated to each other by T4DNA ligase, may be used
CC      for the detection of V. cholerae and enterotoxigenic E.coli
CC      simultaneously in a simple way.
CC      The method has the same sensitivity as the dangerous radio-isotope
CC      detection method.
XX
SQ      Sequence 1148 BP; 399 A; 205 C; 229 G; 315 T; 0 other;

Alignment Scores:
Pred. No.:           1,39e-06              Length:         1148
Score:               15.00                 Matches:          15
Percent Similarity:  100.00%                Conservative:    0
Best Local Similarity: 100.00%             Mismatches:       0
Query Match:        100.00%                 Indels:          0
DB:                  13                     Gaps:            0

SEOL-657079 (1-15) x AAQ23864 (1-1148)

Oy      1 SerLeuArgSerAlaHisLeuAlaGlycInserIleuSergly 15
        |||||
Db      247 AGTTTGAGAAAGTGCCTCACTTAGCAGACGACGTATATTATTCAGGA 251

RESULT 9
ABLA0639 standard; DNA; 1152 BP.
ID      ABLA0639
XX
AC      ABLA0639;
XX
DT      17-JUN-2002 *first entry)
XX
DE      E. coli heat-labile toxin (LT) 5B-SD-1A protein encoding DNA.
XX
KW      LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant;
KW      vaccine; ds.
XX
OS      Escherichia coli.
XX
Key      Location/Qualifiers
FH      1..1152
CDS     /*tag= a
FT      /note= "contains an internal SD sequence"
FT      1..312
FT      /*tag= b
FT      /note= "LT 5B subunit"
FT      313..339
FT      /*tag= c
FT      /note= "SD sequence"
FT      misc_feature 340..1152
FT      /*tag= d
FT      /note= "LT 1A subunit"
XX
XX      JP2002051779-A.
XX
XX      19-FEB-2002.
XX
XX      07-AUG-2000; 2000JP-0238740.
```

```
PR 07-AUG-2000; 2000JP-0238740.
XX
XX (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.
PA (HGET) HIGETA SHOYU KK.
PA (FUJI-) FUJITA GAKUEN.
XX
XX WPI: 2002-299402/34.
DR P-PSDB: ABB07784.
XX
XX Preparation of a protein having 1A5B structure -
PS Claim 2; Page 12; 27/pp; Japanese.
XX
XX The invention relates to a gene encoding a protein having a subunit
CC structure of 1A5B in which the DNA sequence encoding each signal is
CC deleted from the A subunit gene and the B subunit gene and they are
CC combined tandemly in the order of (B subunit gene)-(SD sequence gene)-
CC (A subunit gene). A method is provided for the preparation of a protein
CC having a subunit structure of 1A5B in which the above DNA is connected to
CC a vector expressible in Brevibacillus chosioensis and Brevibacillus
CC chosioensis is transformed by said vector and said transformant is
CC cultured. The protein can be used in the preparation of an adjuvant for
CC vaccine. The present sequence represents the E. coli heat-labile toxin
CC (LT) 5B-SD-1A protein encoding DNA.
XX
SQ Sequence 1152 BP; 396 A; 214 C; 246 G; 296 T; 0 other;

Alignment Scores:
Pred. No.: 1,4e-06 Length: 1152
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

SEQ1-65TO79 (1-15) x ABLA0639 (1-1152)
OY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
Db 622 AGTTTGAGAGAGTGTCTACGACGACAGTCTATATTATTCAGGA 666
RESULT 10
AAS01506
ID AAS01506 standard; DNA; 1508 BP.
XX
XX AAS01506;
XX
XX 29-AUG-2001 (first entry)
XX
XX E. coli heat-labile enterotoxin (LT) mutant LTdel110/112 DNA.
XX
XX Heat-labile enterotoxin; LT; LTS63Y; LTdel110/112; mutant;
KW detoxified and immunologically active protein; ADP-ribosylation; Gs;
KW endotoxin; diarrhoea; ds.
XX
XX Escherichia coli strain K88ac.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 164..1305
XX FT /*tag= a
XX FT /product= "LTdel110/112 mutant protein"
XX FT /transl_except= (pos:932..933,aa:Met)
XX FT /note= "This codon has an apparent 1 nucleotide
XX deletion which alters the reading frame"
XX
XX MO200119998-A1.
XX
XX 22-MAR-2001.
XX
XX 15-SEP-1999; 99MO-KR00555.
XX
XX 15-SEP-1999; 99MO-KR00555.
```

```
XX
XX (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
XX
XX Park EJ, Kim JS, Chang J, Yum J, Chung S;
XX
XX WPI: 2001-281524/29.
DR P-PSDB: AAU00507.
XX
XX New detoxified mutants of Escherichia coli heat-labile enterotoxin
PT useful as vaccine for preventing and treating diarrhoea, and as adjuvant
PT for antibody production -
PS Claim 7; Page 45-46; 48pp; English.
XX
XX The present sequence encodes for Escherichia coli heat-labile
CC enterotoxin (LT) mutant LTdel110/112. LTS63Y (AAU00506) and LTdel110/112
CC are two novel detoxified and immunologically active proteins (LT
CC mutants) derived by site-directed mutagenesis of the A1 subunit of wild
CC type LT. The substitution of Ser to Tyr at position 63 in LTS63Y blocks
CC NAD-binding. Deletion of Glu residues at positions 110 and 112 in
CC LTdel110/112 eliminate the enzymatic activity of LT. The A1 subunit of
CC wild type LT catalyses ADP-ribosylation of Gs, a GTP-binding protein that
CC regulates cAMP levels. The resulting increase in cAMP is the cause of
CC diarrhoea in humans and animals e.g. pigs. The mucosal immunogenicities
CC of mutant heat-labile endotoxins LTS63Y and LTdel110/112 were tested.
CC Groups of mice were immunised with LTS623Y or LTdel110/112. The control
CC faecal antibody titres to LT were determined. The results showed that
CC mice immunised with LTS63Y or LTdel110/112 contained high and
CC comparable level of anti-LT antibodies in sera and faecal extracts
CC compared with those immunised with wild-type LT. The LT mutants are
CC useful as a vaccine for preventing and treating diarrhoea and as an
XX adjuvant for antibody production.
XX
SQ Sequence 1508 BP; 506 A; 277 C; 295 G; 430 T; 0 other;

Alignment Scores:
Pred. No.: 1,79e-06 Length: 1508
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

SEQ1-65TO79 (1-15) x AAS01506 (1-1508)
OY 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
Db 410 AGTTTGAGAGAGTGTCTACGACGACAGTCTATATTATTCAGGA 454
RESULT 10
AAS01506
ID AAS01506 standard; DNA; 1510 BP.
XX
XX AAS01505;
XX
XX 29-AUG-2001 (first entry)
XX
XX E. coli heat-labile enterotoxin (LT) mutant LTS63Y DNA.
XX
XX Heat-labile enterotoxin; LT; LTS63Y; LTdel110/112; mutant;
KW detoxified and immunologically active protein; ADP-ribosylation; Gs;
KW endotoxin; diarrhoea; ds.
XX
XX Escherichia coli strain K88ac.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 164..1311
XX FT /*tag= a
XX FT /product= "LTS63Y mutant protein"
XX FT /transl_except= (pos:938..939,aa:Met)
XX FT /note= "This codon has an apparent 1 nucleotide
XX deletion which alters the reading frame"
XX
```

deletion which alters the reading frame"

```

P.  MO200119998-A1.
XX
XX  22-MAR-2001.
XX
XX  15-SEP-1999; 99MO-KR00555.
XX
XX  15-SEP-1999; 99MO-KR00555.
XX
XX  (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
XX
XX  Park EJ, Kim JS, Chang J, Yun J, Chung S;
XX  WPI; 2001-281524/29.
XX  P-PsDB; AA000506.
XX
XX  New detoxified mutants of Escherichia coli heat-labile enterotoxin
XX  useful as vaccine for preventing and treating diarrhoea, and as adjuvant
XX  for antibody production
XX
XX  Claim 3; Page 41-42; 48pp; English.
XX
XX  The present sequence encodes for Escherichia coli heat-labile
XX  enterotoxin (LT) mutant LT563y. LT563y and LTdel110/112 (AAU00507)
XX  are two novel detoxified and immunologically active proteins (LT
XX  mutants) derived by site-directed mutagenesis of the A1 subunit of wild
XX  type LT. The substitution of Ser to Tyr at position 63 in LT563y blocks
XX  NAD-binding. Deletion of Glu residues at positions 110 and 112 in
XX  LTdel110/112 eliminate the enzymatic activity of LT. The A1 subunit of
XX  wild type LT catalyses ADP-ribosylation of Gs, a GTP-binding protein that
XX  regulates cAMP levels. The resulting increase in cAMP is the cause of
XX  diarrhoea in humans and animals e.g. pigs. The mucosal immunogenicities
XX  of mutant heat-labile enterotoxins LT563y and LTdel110/112 were tested.
XX  Groups of mice were immunised with LT5623y or LTdel110/112. The control
XX  groups received phosphate buffered saline (PBS) alone. The serum and
XX  faecal antibody titres to LT were determined. The results showed that
XX  mice immunised with LT563y or LTdel110/112 contained high and
XX  comparable level of anti-LT antibodies in sera and faecal extracts
XX  compared with those immunised with wild-type LT. The LT mutants are
XX  useful as a vaccine for preventing and treating diarrhoea and as an
XX  adjuvant for antibody production.
XX
XX  Sequence 1514 BP; 510 A; 277 C; 298 G; 429 T; 0 other;
XX
XX  Alignment Scores:
XX  Pred. No.: 1.79e-06 Length: 1514
XX  Score: 15.00 Matches: 15
XX  Percent Similarity: 100.00% Conservative: 0
XX  Best Local Similarity: 100.00% Mismatches: 0
XX  Query Match: 100.00% Indels: 0
XX  DB: 22 Gaps: 0
XX
XX  SEQ1-65to79 (1-15) x AAS01505 (1-1514)
XX
XX  Oy 1 SerLeuArgSerAlaHisLeuAlaGlyClnSerIleLeuSerGly 15
XX  |||||||
XX  Db 410 AGTTTGAGAAAGTGCCTCACTTAGCAGACAGCTCTATATTATCAGGA 454
XX
XX  RESULT 12
XX  ID AAS51140 standard; DNA: 40 BP.
XX
XX  AC AAS51140;
XX
XX  26-SEP-2000 (first entry)
XX
XX  Oligomer SLT-A Rev-14 for LT-A cDNA synthesis.
XX
XX  Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
XX  adjuvant; anti-bacterial; ss.
XX
XX  Escherichia coli.

```

```

OS Synthetic.
XX
XX  PN MO200037609-A2.
XX
XX  29-JUN-2000.
XX
XX  22-DEC-1999; 99MO-US30747.
XX
XX  22-DEC-1998; 98US-0113507.
XX
XX  (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX  (MASO/) MASON H S.
XX  (ARNT/) ARNTZEN C J.
XX
XX  PI Mason HS, Arntzen CJ;
XX
XX  WPI; 2000-442653/38.
XX
XX  New polynucleotides encoding LT-A or CT-A polypeptides for the
XX  transformation of plant cells, useful in immunogenic compositions to
XX  elicit immune responses in animals
XX
XX  Example 1; Fig 2; 103pp; English.
XX
XX  Oligonucleotide 40-mers (AAS51107-26) spanning the entire synthetic
XX  Escherichia coli heat-labile toxin (LT) A subunit (LT-A) coding sequence
XX  and a set of complementary 40-mers (AAS51127-46) that centre on the
XX  junctions of the coding oligomers allowing 20 bp overlaps were obtained
XX  commercially and used to construct the synthetic LT-A cDNA. The sequence
XX  contains plant-preferred codons and eliminates sequence motifs
XX  associated with spurious mRNA processing. A single codon insertion (GTG
XX  encoding valine) was made to accommodate the creation of a NcoI
XX  restriction site around the initiator methionine codon. Novel
XX  polynucleotides encode a mutant LT-A polypeptide or a mutant Vibrio
XX  cholerae cholera toxin (CT) A subunit (CT-A) polypeptide, which have
XX  reduced enzyme activity as compared to the wild-type LT-A or CT-A
XX  polypeptide and where at least one of the codons is altered to a plant
XX  preferred codon. The polynucleotide further comprises a nucleic acid
XX  sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The
XX  polynucleotides are useful for the transformation of plant cells for the
XX  production of transgenic plants to produce edible vaccines, especially
XX  oral vaccines in transgenic plants for the prophylactic or therapeutic
XX  treatment against E. coli or V. cholerae. The mutant polypeptides are
XX  also useful as adjuvants.
XX
XX  SQ Sequence 40 BP; 11 A; 9 C; 13 G; 7 T; 0 other;
XX
XX  Alignment Scores:
XX  Pred. No.: 8.45e-05 Length: 40
XX  Score: 12.00 Matches: 12
XX  Percent Similarity: 100.00% Conservative: 0
XX  Best Local Similarity: 100.00% Mismatches: 0
XX  Query Match: 80.00% Indels: 0
XX  DB: 21 Gaps: 0
XX
XX  SEQ1-65to79 (1-15) x AAS51140 (1-40)
XX
XX  Oy 2 LeuArgSerAlaHisLeuAlaGlyClnSerIleLeu 13
XX  |||||||
XX  Db 38 TTGAGGTCTGCTCACTTAGCAGACAAATCATCCTC 3
XX
XX  RESULT 13
XX  ID AAO51314 standard; DNA: 711 BP.
XX
XX  AC AAO51314;
XX
XX  08-DEC-1993 (first entry)
XX
XX  Encodes Asp-53 E.coli heat labile toxin subunit A.
XX
XX  enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
XX  protomer A; site-directed mutagenesis; reduced toxicity;
XX

```

```
KM ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_difference 153..155
FT /tag= a
FT /note= "wild-type GTC(Val) mutated to GAT(Asp)"
XX
XX
PN W09313202-A.
XX
PD 08-JUL-1993.
XX
PE 30-DEC-1992: 92WO-EP03016.
XX
PR 31-DEC-1991: 91IT-0M13513.
XX
PA (BIOC-) BIOCINE SCLAVO SPA.
XX
PI Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
DR WPI: 1993-227320/28.
XX
DR P-PSDB: AAR38730.
XX
PT Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
CC The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051314 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 235 A; 122 C; 156 G; 198 T; 0 other;

Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0

SEQ1-65TO79 (1-15) x AA051314 (1-711)
OY 1 SerLeuARgSerAlaHisLeuAlaGlyIn 10
DB 190 AGTTTGAGAGTGTCTACTTAGCGAGACAG 219

RESULT 14
AA051315
ID AA051315 standard; DNA: 711 BP.
XX
AC AA051315;
XX
DE 08-DEC-1993 (first entry)
XX
DE Encodes Glu-53 E.coli heat labile toxin subunit A.
XX
KM enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KM protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
```

```
FH Key Location/Qualifiers
FT misc_difference 153..155
FT /tag= a
FT /note= "wild-type GTC(Val) mutated to GAA(Glu)"
XX
XX
PN W09313202-A.
XX
PD 08-JUL-1993.
XX
PE 30-DEC-1992: 92WO-EP03016.
XX
PR 31-DEC-1991: 91IT-0M13513.
XX
PA (BIOC-) BIOCINE SCLAVO SPA.
XX
PI Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
DR WPI: 1993-227320/28.
XX
DR P-PSDB: AAR38731.
XX
PT Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
CC The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051315 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 236 A; 122 C; 156 G; 197 T; 0 other;

Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0

SEQ1-65TO79 (1-15) x AA051315 (1-711)
OY 1 SerLeuARgSerAlaHisLeuAlaGlyIn 10
DB 190 AGTTTGAGAGTGTCTACTTAGCGAGACAG 219

RESULT 15
AA051316
ID AA051316 standard; DNA: 711 BP.
XX
AC AA051316;
XX
DE 08-DEC-1993 (first entry)
XX
DE Encodes Tyr-53 E.coli heat labile toxin subunit A.
XX
KM enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KM protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_difference 153..155
FT /tag= a
FT /note= "wild-type GTC(Val) mutated to TAC(Tyr)"
XX
```

```

XX MO9313202-A.
PN
XX
XX 08-JUL-1993.
PD
XX
XX 30-DEC-1992; 92WO-EP03016.
PF
XX
XX 31-DEC-1991; 91IT-0M13513.
PR
XX
XX (BIOC-) BIOGINE SCLAVO SPA.
PA
XX
XX Domenighini M, Hol W, Piazza M, Rappuoli R;
PI
XX
XX WPI: 1993-227320/28.
DR
XX
XX P-PSDB: AAR38732.
DR
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
PT
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
PS
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al., J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051316 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
CC
XX
SQ Sequence 711 BP; 235 A; 123 C; 155 G; 198 T; 0 other;

Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0

SEQ1-65TO79 (1-15) x AA051316 (1-711)
OY 1 SerLeuArgSerAlaHisLeuAlaGlyIn 10
Db 190 AGTTTGAGAGAGTCTCAGTACGAGACAG 219
RESULT 16
AA051317
ID AA051317 standard; DNA; 711 BP.
XX
XX AA051317;
AC
XX
XX 08-DEC-1993 (first entry)
DT
XX
XX Encodes Lys-63 E.coli heat labile toxin subunit A.
DE
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity; ss.
XX
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FH misc_difference 184..186
FT /*tag= a
FT /note= "wild-type rct(Ser) mutated to AAG(Lys)"
XX
XX MO9313202-A.
PN
XX
XX 08-JUL-1993.
PD

```

```

XX
XX 30-DEC-1992; 92WO-EP03016.
PF
XX
XX 31-DEC-1991; 91IT-0M13513.
PR
XX
XX (BIOC-) BIOGINE SCLAVO SPA.
PA
XX
XX Domenighini M, Hol W, Piazza M, Rappuoli R;
PI
XX
XX WPI: 1993-227320/28.
DR
XX
XX P-PSDB: AAR4016.
DR
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
PT
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
PS
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al., J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051317 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
CC
XX
SQ Sequence 711 BP; 236 A; 122 C; 157 G; 196 T; 0 other;

Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0

SEQ1-65TO79 (1-15) x AA051317 (1-711)
OY 1 SerLeuArgSerAlaHisLeuAlaGlyIn 10
Db 190 AGTTTGAGAGAGTCTCAGTACGAGACAG 219
RESULT 17
AA051318
ID AA051318 standard; DNA; 711 BP.
XX
XX AA051318;
AC
XX
XX 08-DEC-1993 (first entry)
DT
XX
XX Encodes Lys-97 E.coli heat labile toxin subunit A.
DE
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity; ss.
XX
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FH misc_difference 277..279
FT /*tag= a
FT /note= "wild-type GGT(Val) mutated to AAG(Lys)"
XX
XX MO9313202-A.
PN
XX
XX 08-JUL-1993.
PD
XX
XX 30-DEC-1992; 92WO-EP03016.
PF
XX
XX 31-DEC-1991; 91IT-0M13513.
PR

```

```

XX (BIOC-) BIOCINE SCLAVO SPA.
PA
XX
XX Domenighini M, Hol W, Pizza M, Rappuoli R;
PI
XX WPI: 1993-227320/28.
DR
XX P-PSDB: AAR44017.
DR
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
PS
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051318 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 236 A; 123 C; 156 G; 196 T; 0 other;

Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: Gaps: 0

SEQ1-65T079 (1-15) x AA051318 (1-711)

OY 1 SerLeuArgSerAlaHisLeuAlaGlyGln 10
DB 190 AGTTTGAGAGAGTGCTCAGCTTAGCAGACAG 219

RESULT 18
AA051319
ID AA051319 standard; DNA; 711 BP.
XX
AC AA051319;
XX
DT 08-DEC-1993 (first entry)
XX
DE Encodes Tyr-97 E.coli heat labile toxin subunit A.
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KM ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
XX Key Location/Qualifiers
FH misc_difference 277..279
FT /*tag= a
FT /note= "wild-type GTR(Val) mutated to TAC(Tyr)"
FT
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92MO-EP03016.
XX
XX 31-DEC-1991; 91IT-0MI3513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Pizza M, Rappuoli R;
PI

```

```

XX WPI: 1993-227320/28.
DR
XX P-PSDB: AAR44018.
DR
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
PS
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051319 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 235 A; 124 C; 155 G; 197 T; 0 other;

Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: Gaps: 0

SEQ1-65T079 (1-15) x AA051319 (1-711)

OY 1 SerLeuArgSerAlaHisLeuAlaGlyGln 10
DB 190 AGTTTGAGAGAGTGCTCAGCTTAGCAGACAG 219

RESULT 19
AA051320
ID AA051320 standard; DNA; 711 BP.
XX
AC AA051320;
XX
DT 08-DEC-1993 (first entry)
XX
DE Encodes Glu-107 E.coli heat labile toxin subunit A.
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KM ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
XX Key Location/Qualifiers
FH misc_difference 307..309
FT /*tag= a
FT /note= "wild-type CAC(His) mutated to GAG(Glu)"
FT
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92MO-EP03016.
XX
XX 31-DEC-1991; 91IT-0MI3513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Pizza M, Rappuoli R;
PI
XX WPI: 1993-227320/28.
DR
XX P-PSDB: AAR44019.
XX

```


PT Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX
PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al., J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051320 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 234 A; 121 C; 158 G; 198 T; 0 other;
SO
Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0
SEQ1-65TO79 (1-15) x AA051320 (1-711)
QY 1 SerLeuArgSerAlaHisLeuAlaGlyGln 10
ID 190 AGTTTGAGAAAGTCTCTACTTAGCAGGACAG 219
DB 190 AGTTTGAGAAAGTCTCTACTTAGCAGGACAG 219
XX
RESULT 20
AA051321
ID AA051321 standard; DNA: 711 BP.
XX
AC AA051321:
XX
DT 08-DEC-1993 (first entry)
XX
DE Encodes Lys-104 E.coli heat labile toxin subunit A.
XX
KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KM protomer A; site-directed mutagenesis; reduced toxicity;
XX ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_difference 298..300
FT /*tag= a
FT /note= "wild-type TAC(Tyr) mutated to AAG(Lys)"
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92WO-EP03016.
XX
XX 31-DEC-1991; 91IT-0MI3513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Piazza M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX P-PSDB; AAR44020.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX

PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al., J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-Q51326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051321 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 235 A; 122 C; 157 G; 197 T; 0 other;
SO
Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0
SEQ1-65TO79 (1-15) x AA051321 (1-711)
QY 1 SerLeuArgSerAlaHisLeuAlaGlyGln 10
ID 190 AGTTTGAGAAAGTCTCTACTTAGCAGGACAG 219
DB 190 AGTTTGAGAAAGTCTCTACTTAGCAGGACAG 219
XX
RESULT 21
AA051322
ID AA051322 standard; DNA: 711 BP.
XX
AC AA051322:
XX
DT 08-DEC-1993 (first entry)
XX
DE Encodes Asp-104 E.coli heat labile toxin subunit A.
XX
KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KM protomer A; site-directed mutagenesis; reduced toxicity;
XX ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_difference 298..300
FT /*tag= a
FT /note= "wild-type TAC(Tyr) mutated to GAT(Asp)"
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92WO-EP03016.
XX
XX 31-DEC-1991; 91IT-0MI3513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Piazza M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX P-PSDB; AAR44021.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC

CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-051326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051322 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 234 A; 122 C; 157 G; 198 T; 0 other;
Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0
SEQ1-65to79 (1-15) x AA051322 (1-711)
QY 1 SerLeuArgSerAlaHisLeuAlaGlyIn 10
DB 190 AGTTTGAGAGTGCCTCAGTTCAGACAGACAG 219
RESULT 22
AA051323
ID AA051323 standard; DNA; 711 BP.
XX
AC AA051323;
XX
DT 08-DEC-1993 (first entry)
XX
DE Encodes Ser-104 E.coli heat labile toxin subunit A.
XX
KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_difference 298..300
FT /*tag= a
FT /note= "wild-type TAC(Tyr) mutated to AGT(Ser)"
XX
PN WO9313202-A.
XX
PD 08-JUL-1993.
XX
PF 30-DEC-1992; 92MO-EP03016.
XX
PR 31-DEC-1991; 91IT-OMI3513.
XX
PA (BIOC-) BIOCINE SCLAVO SPA.
XX
PI Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
DR WPI: 1993-227320/28.
DR P-PSDB; AAR44022.
XX
PT Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-051326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051322 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
CC immunogenic, detoxified LT-A proteins and their use in vaccines to

CC protect against enterotoxigenic E.coli. Sequence AA051323 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 711 BP; 234 A; 122 C; 157 G; 198 T; 0 other;
Alignment Scores:
Pred. No.: 0.14 Length: 711
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0
SEQ1-65to79 (1-15) x AA051323 (1-711)
QY 1 SerLeuArgSerAlaHisLeuAlaGlyIn 10
DB 190 AGTTTGAGAGTGCCTCAGTTCAGACAGACAG 219
RESULT 23
AA051324
ID AA051324 standard; DNA; 711 BP.
XX
AC AA051324;
XX
DT 08-DEC-1993 (first entry)
XX
DE Encodes Ser-106 E.coli heat labile toxin subunit A.
XX
KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT misc_difference 304..306
FT /*tag= a
FT /note= "wild-type CCT(Pro) mutated to AGT(Ser)"
XX
PN WO9313202-A.
XX
PD 08-JUL-1993.
XX
PF 30-DEC-1992; 92MO-EP03016.
XX
PR 31-DEC-1991; 91IT-OMI3513.
XX
PA (BIOC-) BIOCINE SCLAVO SPA.
XX
PI Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
DR WPI: 1993-227320/28.
DR P-PSDB; AAR44023.
XX
PT Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AA051314-051326). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. Sequence AA051324 is a
CC combination of the wild-type coding sequence and the mutagenic
CC primer sequence used to introduce the preferred mutation. (Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
CC acid numbering is based on the cholera toxin A subunit sequence).

```

XX      SQ      Sequence 711 BP; 235 A; 122 C; 157 G; 197 T; 0 other;
Alignment Scores:
Pred. No.:      0.14      Length:      711
Score:          10.00     Matches:      10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    66.67%   Indels:      0
DB:            14       Gaps:        0

SEQ1-65TO79 (1-15) x AAQ51324 (1-711)

QY      1 SerLeuArgSerAlaHisLeuAlaGlyGln 10
DB      190 AGTTGAGAACTGCTCACTTAGCAGACAG 219

RESULT 24
AAQ51325
ID      AAQ51325 standard; DNA; 711 BP.
XX
AC      AAQ51325;
XX
DT      08-DEC-1993 (first entry)
XX
DE      Encodes Glu-114 E.coli heat labile toxin subunit A.
XX
KW      enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW      protomer A; site-directed mutagenesis; reduced toxicity;
KW      ADP-ribosyltransferase activity; ss.
OS      Escherichia coli.
XX
FH      Key Location/Qualifiers
FT      misc_difference 328..330
FT      /*tag= a
FT      /note= "wild-type TCT(Ser) mutated to GAA(Glu)"
XX
PN      W09313202-A.
XX
PD      08-JUL-1993.
XX
PF      30-DEC-1992; 92WO-EP03016.
XX
PR      31-DEC-1991; 91IT-0MT3513.
XX
PA      (BIOC-) BIOCINE SCLAVO SPA.
XX
PI      Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
DR      WPI: 1993-227320/28.
DR      P-PSDB: AAR44024.
XX
PT      Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT      - useful as vaccines against infection by Vibrio cholerae and
PT      enterotoxin producing Escherichia coli
XX
PS      Claim 3; Fig 2 and Page 46; 60pp; English.
XX
CC      The wild-type sequence coding for the A subunit of the heat labile
CC      toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC      et al, J.Biol. Chem., 259, 5037-5044 - see AAQ42768) was subjected to
CC      site-directed mutagenesis. Certain mutations were found to reduce
CC      toxicity (see AAQ51314-Q51326). The invention relates to
CC      immunogenic, detoxified LT-A proteins and their use in vaccines to
CC      protect against enterotoxigenic E.coli. Sequence AAQ51325 is a
CC      combination of the wild-type coding sequence and the mutagenic
CC      primer sequence used to introduce the preferred mutation. (Amino
CC      acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ      Sequence 711 BP; 236 A; 122 C; 157 G; 196 T; 0 other;
Alignment Scores:

```

```

Pred. No.:      0.14      Length:      711
Score:          10.00     Matches:      10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    66.67%   Indels:      0
DB:            14       Gaps:        0

SEQ1-65TO79 (1-15) x AAQ51325 (1-711)

QY      1 SerLeuArgSerAlaHisLeuAlaGlyGln 10
DB      190 AGTTGAGAACTGCTCACTTAGCAGACAG 219

RESULT 25
AAQ51326
ID      AAQ51326 standard; DNA; 711 BP.
XX
AC      AAQ51326;
XX
DT      08-DEC-1993 (first entry)
XX
DE      Encodes Lys-114 E.coli heat labile toxin subunit A.
XX
KW      enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW      protomer A; site-directed mutagenesis; reduced toxicity;
KW      ADP-ribosyltransferase activity; ss.
OS      Escherichia coli.
XX
FH      Key Location/Qualifiers
FT      misc_difference 328..330
FT      /*tag= a
FT      /note= "wild-type TCT(Ser) mutated to AAA(Lys)"
XX
PN      W09313202-A.
XX
PD      08-JUL-1993.
XX
PF      30-DEC-1992; 92WO-EP03016.
XX
PR      31-DEC-1991; 91IT-0MT3513.
XX
PA      (BIOC-) BIOCINE SCLAVO SPA.
XX
PI      Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
DR      WPI: 1993-227320/28.
DR      P-PSDB: AAR44025.
XX
PT      Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT      - useful as vaccines against infection by Vibrio cholerae and
PT      enterotoxin producing Escherichia coli
XX
PS      Claim 3; Fig 2 and Page 46; 60pp; English.
XX
CC      The wild-type sequence coding for the A subunit of the heat labile
CC      toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC      et al, J.Biol. Chem., 259, 5037-5044 - see AAQ42768) was subjected to
CC      site-directed mutagenesis. Certain mutations were found to reduce
CC      toxicity (see AAQ51314-Q51326). The invention relates to
CC      immunogenic, detoxified LT-A proteins and their use in vaccines to
CC      protect against enterotoxigenic E.coli. Sequence AAQ51326 is a
CC      combination of the wild-type coding sequence and the mutagenic
CC      primer sequence used to introduce the preferred mutation. (Amino
CC      acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ      Sequence 711 BP; 237 A; 122 C; 156 G; 196 T; 0 other;
Alignment Scores:
Pred. No.:      0.14      Length:      711
Score:          10.00     Matches:      10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 66.67% Indels: 0
DB: 14 Gaps: 0

SEQ1-65to79 (1-15) x AAQ51326 (1-711)

OY 1 SerLeuArgSerAlaHisLeuAlaGlycin 10
DB 190 AGTTTGAGAGTCTCCTTAGCAGACAG 219

RESULT 26

AAQ42768
ID AAQ42768 standard; DNA; 711 BP.

AC AAQ42768;

DT 08-DEC-1993 (first entry)

DE E.coli heat labile toxin subunit A coding sequence.

XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;

KW protomer A; site-directed mutagenesis; reduced toxicity;

KM ADP-ribosyltransferase activity; ss.

XX Escherichia coli.

XX Key Location/Qualifiers

FT mat_peptide 1..708 /tag= a

FT /note= "LT-A"

XX WO9313202-A.

XX 08-JUL-1993.

XX 30-DEC-1992; 92MO-EP03016.

XX 31-DEC-1991; 91IT-OM13513.

XX (BIOC-) BIOCINE SCLAVO SPA.

XX Domenighini M, Hol W, Pizze M, Rappuoli R;

XX WPI; 1993-227320/28.

XX P-PSDB; AAR38728.

XX Immunogenic detoxified mutant cholera toxin and heat labile toxin

PT - useful as vaccines against infection by Vibrio cholerae and

PT enterotoxin producing Escherichia coli

XX Disclosure; Fig 2; 60pp; English.

XX This sequence encodes the A subunit of the heat labile toxin (LT-A)

CC of a strain of E.coli known to affect humans. The sequence was

CC published by Yamamoto et al, J.Biol. Chem., 259, 5037-5044.

CC Mutations at selected positions within this sequence have been found

CC to reduce toxicity (see AAQ51314-Q51326). The invention relates to

CC immunogenic, detoxified LT-A proteins and their use in vaccines to

XX protect against enterotoxigenic E.coli.

XX Sequence 711 BP; 234 A; 123 C; 156 G; 198 T; 0 other;

SO Alignment Scores:

Pred. No.: 0.14 Length: 711

Score: 10.00 Matches: 10

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 66.67% Indels: 0

DB: 14 Gaps: 0

SEQ1-65to79 (1-15) x AAQ42768 (1-711)

OY 1 SerLeuArgSerAlaHisLeuAlaGlycin 10

|||||

DB 190 AGTTTGAGAGTCTCCTTAGCAGACAG 219

RESULT 27

AAV81595

ID AAV81595 standard; DNA; 711 BP.

AC AAV81595;

XX 11-MAY-1999 (first entry)

DE E. coli heat labile toxin gene.

XX A subunit; heat labile toxin; ADP-ribosylation; mutant; detoxification;

KW parenteral adjuvant; antigen; immunisation; humoral response;

KW cell-mediated immune response; virus; bacterium; parasite; fungus;

KW tumour; allergen; pathogen; AIDS; autoimmune disease; cancer; antibody;

KW systemic lupus erythematosus; Alzheimer's disease; diagnosis; ss.

XX Escherichia coli.

XX WO9842375-A1.

XX 01-OCT-1998.

XX 19-MAR-1998; 98WO-US05454.

XX 18-MAR-1998; 98US-0044696.

XX 21-MAR-1997; 97US-0041227.

XX (CHIR) CHIRON CORP.

XX Barchfeld G, Del Giudice G, Rappuoli R;

XX WPI; 1999-070064/06.

XX P-PSDB; AAW67712.

XX Detoxified mutants of bacterial ADP-ribosylating toxins as

PT parenteral adjuvants - useful to enhance humoral and cell-mediated

PT immune responses in vertebrates when administered with selected

PT antigen e.g. in disease treatment

XX Disclosure; Fig 1A-B; 51pp; English.

XX This sequence corresponds to the coding region for the A subunit of the

CC E. coli heat labile toxin, an example of a bacterial ADP-ribosylating

CC toxin. A mutant detoxified form of this protein is used in a parenteral

CC adjuvant composition, which comprises the detoxified protein, at least

CC one selected antigen and optionally a pharmaceutically acceptable

CC (optionally topical) vehicle. The adjuvant composition can be

CC administered parenterally in conjunction with at least one antigen in

CC methods to immunise vertebrate subjects. The adjuvant has the ability

CC to enhance the humoral and cell-mediated immune responses elicited by

CC the antigen (e.g. by making the antigen more strongly immunogenic or

CC necessitating fewer/lower antigen doses). It can be administered

CC prior/subsequent to the antigen, and is preferably administered within

CC a short space of time to the same site; it can also be administered in

CC isolation from antigens as a boost following systemic or mucosal antigen

CC administration. Most preferably, the adjuvant is co-administered with

CC the antigen in the compositions and a pharmaceutically acceptable

CC carrier. The antigen may be derived from viruses, bacteria, parasites

CC and fungi or may be tumour antigens, self-antigens and allergens. The

CC compositions are therefore useful in the treatment and prevention of

CC e.g. viral diseases, allergic manifestations, diseases caused by

CC pathogens (e.g. bacteria or parasites), AIDS, autoimmune diseases

CC (e.g. Systemic Lupus Erythematosus), Alzheimer's disease and cancers.

CC The adjuvant can also be used to prepare antibodies against selected

CC antigen(s), useful e.g. for diagnostic purposes or for antigen

CC purification.

XX Sequence 711 BP; 234 A; 123 C; 156 G; 198 T; 0 other;

SO Alignment Scores:

Pred. No.: 0.14 Length: 711

Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 66.67% Indels: 0
 DB: 20 Gaps: 0

SEQ1-65to79 (1-15) x AAV81595 (1-711)

OY 1 SerleuArgSerAlaHisIleuAlaGlyIle 10
 |||||
 Db 190 AGTTGAGAGTCTCCTCACTTAGCAGACAG 219

RESULT 28

AAA5114
 ID AAA5114 standard; DNA: 40 BP.

XX AAA5114:

XX 26-SEP-2000 (first entry)

XX Oligomer sLT-A For-8 for LT-A cDNA synthesis.

XX Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
 KW adjuvant; anti-bacterial; ss.

XX Escherichia coli.
 OS Synthetic.

XX WO200037609-A2.

XX 29-JUN-2000.

XX 22-DEC-1999; 99WO-US30747.

XX 22-DEC-1998; 98US-0113507.

XX (BOYC-) BOYCE THOMPSON INST PLANT RES.

PA (MASO/) MASON H S.

PA (ARNT/) ARNTZEN C J.

PI Mason HS, Arntzen CJ;

DR WPI; 2000-442653/38.

PT New polynucleotides encoding LT-A or CT-A polypeptides for the
 PT transformation of plant cells; useful in immunogenic compositions to
 PT elicit immune responses in animals

PS Example 1; Fig 2; 103pp; English.

XX Oligonucleotide 40-mers (AAA5107-26) spanning the entire synthetic
 CC Escherichia coli heat-labile toxin (LT) A subunit (LT-A) coding sequence
 CC and a set of complementary 40-mers (AAA5117-46) that centre on the
 CC junctions of the coding oligomers allowing 20 bp overlaps were obtained
 CC commercially and used to construct the synthetic LT-A cDNA. The sequence
 CC contains plant-preferred codons and eliminates sequence motifs
 CC associated with spurious mRNA processing. A single codon insertion (GTG
 CC encoding valine) was made to accommodate the creation of a NcoI
 CC restriction site around the initiator methionine codon. Novel
 CC polynucleotides encode a mutant LT-A polypeptide or a mutant Vibrio
 CC cholerae cholera toxin (CT) A subunit (CT-A) polypeptide, which have
 CC reduced enzyme activity as compared to the wild-type LT-A or CT-A
 CC polypeptide and where at least one of the codons is altered to a plant
 CC preferred codon. The polynucleotide further comprises a nucleic acid
 CC sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The
 CC polynucleotides are useful for the transformation of plant cells for the
 CC production of transgenic plants to produce edible vaccines, especially
 CC oral vaccines in transgenic plants for the prophylactic or therapeutic
 CC treatment against E. coli or V. cholerae. The mutant polypeptides are
 CC also useful as adjuvants.

XX Sequence 40 BP; 12 A; 15 C; 5 G; 8 T; 0 other;

Alignment Scores:
 Pred. No.: 1.2 Length: 40
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 21 Gaps: 0

SEQ1-65to79 (1-15) x AAA5114 (1-40)

OY 8 AlAGlyGlnSerIleLeuSergly 15
 |||||
 Db 1 GCAGGACATCATCCTCTCAGGA 24

Search completed: January 28, 2003, 10:56:27
 Job time : 345.333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:36:45 ; Search time 1270.67 Seconds
(without alignments)
191.185 Million cell updates/sec

Title: SEQ1-65to79
Perfect score: 15
Sequence: 1 SIRSARHLACGSIISLG 15

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 8

Total number of hits satisfying chosen parameters: 21

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgm2_1/USPFO.spool/BORIN683/rnat_23012003_130141_7880/app.query.fasta_1.597
-DB=EST -QPM=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=8 -ALIGN=50 -MODE=LOCAL -OUTPMT=pto
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=BORIN682.ecgn.1.1.1349 -rnat_23012003_130141_7880 -NCPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-MARK_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estc2:*
11: gb_hlc:*
12: gb_estc3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_liv:*
20: em_gss_pln:*
21: em_gss_vtl:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	53.3	263	17	A2788524	A2788524 2M0035K08
2	8	53.3	418	9	A1347100	A1347100 qp55a12.x
3	8	53.3	424	17	A0786225	A0786225 HS_3137_A
4	8	53.3	454	12	BE744213	BE744213 601577186
5	8	53.3	490	17	A0209383	A0209383 HS_3240_A
6	8	53.3	503	12	BF443144	BF443144 260525_MA
7	8	53.3	535	13	B1343334	B1343334 371448_MA
8	8	53.3	615	17	BH329082	BH329082 CH230-105
9	8	53.3	646	12	BG824614	BG824614 602728548
10	8	53.3	648	17	A2403745	A2403745 1M0171010
11	8	53.3	706	13	B1090511	B1090511 602855611
12	8	53.3	725	17	BH271552	BH271552 CH230-34M
13	8	53.3	725	17	BH351398	BH351398 CH230-175
14	8	53.3	768	13	B1658656	B1658656 603283613
15	8	53.3	790	17	CNS03A1W	AL234653 Tetraodon
16	8	53.3	816	17	A0745347	A0745347 HS_2276_A
17	8	53.3	829	17	CNS03WYZ	AL264336 Tetraodon
18	8	53.3	870	12	BF120183	BF120183 601756489
19	8	53.3	884	12	BF575766	BF575766 602135396
20	8	53.3	967	17	CNS03N3Z	AL251576 Tetraodon
21	8	53.3	1877	11	BC019121	BC019121 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS A2788524 263 bp DNA linear GSS 16-FEB-2001
DEFINITION c10ne035K08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
ACCESSION UUGC2M0035K08 R. DNA sequence.
VERSION A2788524.1 GI:12928413
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Plasmid Inserts
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std error: 0.00
Plate: 0035 row: K column: 08
Seq primer: CACACAGAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 263.
Location/Qualifiers
1. 263
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0035K08"

FEATURES
source

/clone_1lb="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD2mV; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/notes/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1473211419b1AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 47 a 81 c 78 g 57 t
 ORIGIN

Alignment Scores:
 Pred. No.: 180 Length: 263
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 17 Gaps: 0

SEQ1-65T079 (1-15) x AZ788524 (1-263)

QY 5 A1aH1sLeuA1aG1yInser1le 12
 Db 161 GCTCACCTGGCGGACAGACATT 184

RESULT 2
 A1347100/c 418 bp mRNA linear EST 30-DEC-1998
 LOCUS 9p55a12.x1 NCI-CGAP_C08 Homo sapiens CDNA IMAGE:1926910 3'
 DEFINITION similar to TR:015498 015498 SNARE PROTEIN XRT6. ;, mRNA sequence.
 ACCESSION A1347100
 VERSION A1347100.1 GI:4084306
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE 1 (bases 1 to 418)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgepbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmerit-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

FEATURES
 source
 Trace considered overall poor quality
 Seq primer: -40UP from gibco
 High quality sequence stop: 1.
 Location/Qualifiers
 1..418
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="IMAGE:1926910"
 /clone_1lb="NCI CGAP_C08"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 110 a 109 c 102 g 97 t
 ORIGIN

Alignment Scores:
 Pred. No.: 316 Length: 418
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 9 Gaps: 0

SEQ1-65T079 (1-15) x A1347100 (1-418)

QY 7 LeuA1aG1yInser1leUser 14
 Db 336 CTTGCGGCGCAGTCATCTGTG 313

RESULT 3
 A0786225/c 424 bp DNA linear GSS 03-AUG-1999
 LOCUS HS_3137_A1_A11_T7C CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3137 Col=21 Row=A, DNA sequence.
 ACCESSION A0786225
 VERSION A0786225.1 GI:5693849
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE 1 (bases 1 to 424)
 Mahatras G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 JOURNAL COMMENT Contact: Mahatras GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 424.
 Location/Qualifiers
 1..424
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3137 Col=21 Row=A"
 /clone_1lb="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 120 a 76 c 83 g 144 t 1 others


```

ORIGIN
Alignment Scores:
Pred. No.: 322 Length: 424
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 17 Gaps: 0

SEQ1-65T079 (1-15) x AQ786225 (1-424)

QY 7 LeuAlaGlycInSerIleLeuSer 14
|||||
Db 134 CTGCAGCTCAGTCTACTTCA 111

RESULT 4
BE744213 454 bp mRNA linear EST 15-SEP-2000
LOCUS BE744213 601577186F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838221 5',
DEFINITION mRNA sequence.
ACCESSION BE744213
VERSION BE744213.1 GI:10158205
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 454)
REFERENCE NIH-MGC http://mgi.ncl.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bts-remail.nih.gov
Tissue Procurement: DCTP/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: L1CM524 row: k column: 22
High quality sequence stop: 303.
Location/Qualifiers
1..454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3838221"
/clone_1lb="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pOTB7; Site: 1: XhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT 89 a 150 c 132 g 83 t
ORIGIN
Alignment Scores:
Pred. No.: 350 Length: 454
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 12 Gaps: 0

SEQ1-65T079 (1-15) x BE744213 (1-454)

QY 3 ArgSerAlaHisLeuAlaGlycIn 10
|||||
Db 134 CTGCAGCTCAGTCTACTTCA 111

RESULT 5
AQ209383 490 bp DNA linear GSS 18-SEP-1998
LOCUS BE443144/C 503 bp mRNA linear EST 01-DEC-2000
DEFINITION BE443144 260525 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE443144
VERSION BE443144.1 GI:11503236
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 503)
REFERENCE Fahrénkrug, S.C., Frekling, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
```

TITLE and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPJ
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithem@mail.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCGACGACGACG

Plate: 92 row: P column: 13

Seq primer: ATTTAGTGACACTATAG.

FEATURES
source
1..503
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site.1: NotI; Site.2: SalI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 95 a 143 c 148 g 116 t 1 others
ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
8.00	100.00%	8	Conservative: 0
100.00%	100.00%	0	Mismatches: 0
53.33%	53.33%	0	Indels: 0
12			Gaps: 0

SEQ1-65to79 (1-15) x BF443144 (1-503)

Qy 1 SerleuArgSerAlaHisLeuAla 8
|||||
DB 228 AGCTGTGAGAGCGCTCACCCTGGCG 205

RESULT 7

BI343334/c

LOCUS BI343334 535 bp mRNA linear EST 30-JUL-2001

DEFINITION 371448 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BI343334

VERSION BI343334.1 GI:15036623

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 535)

Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,

Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,

and Keeler,J.W.

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

Unpublished (2000)

Contact: Smith TPJ

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smithem@mail.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCGACGACGACG
Plate: 116 row: E column: 11
Seq primer: ATTTAGTGACACTATAG.

FEATURES

source

1..535

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 2P1G"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site.1: NotI; Site.2: SalI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 97 a 163 c 144 g 131 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
8.00	100.00%	8	Conservative: 0
100.00%	100.00%	0	Mismatches: 0
53.33%	53.33%	0	Indels: 0
13			Gaps: 0

SEQ1-65to79 (1-15) x BI343334 (1-535)

Qy 1 SerleuArgSerAlaHisLeuAla 8
|||||
DB 56 AGCTGTGAGAGCGCTCACCCTGGCG 33

RESULT 8

BH329082

LOCUS BH329082 615 bp DNA linear GSS 03-DEC-2001

DEFINITION CH230-105P11.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone

CH230-105P11, DNA sequence.

ACCESSION BH329082

VERSION BH329082.1 GI:17259796

KEYWORDS GSS.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 615)

Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn

,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de

Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other GSSs: CH230-105P11.TVB

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong (pdjong@mail.cbo.org).

Clones may be purchased from BACPAC Resources

(<http://www.chori.org/bacpac/orering/information.htm>). BAC end

page: http://www.tigr.org/tldb/bac-ends/rat/bac-end_intro.html

Plate: 105 row: P column: 11

Seq primer: SP6

Class: BAC ends.

FEATURES
source
1..615
/organism="Rattus norvegicus"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-105P11"

```

/clone.lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNhsd/MCW) BAC library produced by
Pleier de Jong"
BASE COUNT      88 a      171 c      170 g      186 t
ORIGIN

Alignment Scores:
Pred. No.:      507      Length:      615
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      53.33%      Indels:      0
DB:              17      Gaps:      0

SEQ1-65TO79 (1-15) x BG824614 (1-615)
OY      1 SerleuArgSerAlaHisLeuAla 8
Db      413 AGCCTCAGGAGTCCACCTGGCC 436

RESULT 9
BG824614/c      646 bp      mRNA      linear      EST 22-MAY-2001
LOCUS      602728548f1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4868195 5',
DEFINITION      mRNA sequence.
ACCESSION      BG824614
VERSION      BG824614.1 GI:14172201
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
1 (bases 1 to 646)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LDCM1737 row: 0 column: 12
High quality sequence stop: 639.
Location/Qualifiers
1..646
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="IMAGE:4868195"
/clone.lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      158 a      172 c      168 g      148 t
ORIGIN

Alignment Scores:
Pred. No.:      538      Length:      646
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0

```

```

Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      53.33%      Indels: 0
DB:              12      Gaps: 0

SEQ1-65TO79 (1-15) x BG824614 (1-646)
OY      7 LeuAlaGlyGlnSerIleLeuSer 14
Db      468 CTGGCAGCAGCAGTATCTTGTC 445

RESULT 10
A2403745      648 bp      DNA      linear      GSS 03-OCT-2000
LOCUS      1M0171010R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION      clone UUGC1M0171010 R, DNA sequence.
ACCESSION      A2403745
VERSION      A2403745.1 GI:10527758
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 648)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0171 row: 0 column: 10
Seq primer: CACACGAGAAACGCTATGACC
Class: plasmid ends
High quality sequence stop: 648.
Location/Qualifiers
1..648
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone.lib="UUGC1M0171010"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/nases/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (91473211419b1AF129072.1), a copy number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      125 a      203 c      129 g      191 t
ORIGIN

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
8.00	100.00%	648	8
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	53.33%	Mismatches:	0
DB:	17	Indels:	0
		Gaps:	0

SEQ1-65to79 (1-15) x AZ403745 (1-648)

OY 4 SerAlaHisLeuAlaGlyClnSer 11
|||||

Db 200 TCTGCTCATCTTGCAGCACAAGT 223

RESULT 11

LOCUS BI090511/c 706 bp mRNA linear EST 20-JUN-2001
DEFINITION 602855611F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996971 5',
mRNA sequence.
ACCESSION BI090511
VERSION BI090511.1 GI:14508841
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 706)
NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM1023 row: e column: 04
High quality sequence stop: 704.

FEATURES
Location/Qualifiers
source 1..706
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4996971"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SF0RT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

BASE COUNT 160 a 210 c 191 g 145 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
8.00	100.00%	706	8
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	53.33%	Mismatches:	0
DB:	13	Indels:	0
		Gaps:	0

SEQ1-65to79 (1-15) x BI090511 (1-706)

OY 2 LeuArgSerAlaHisLeuAlaGly 9
|||||

Db 639 TTGAGCTCAGCACAATTGGCAGGT 616

RESULT 12

BH271552

LOCUS BH271552 725 bp DNA linear GSS 30-NOV-2001
DEFINITION CH230-34M7.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-34M7, DNA sequence.
ACCESSION BH271552
VERSION BH271552.1 GI:17183954
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 725)
Zhao,S., Shetty,U., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcORI segment
Unpublished (1999)
Other_GSSs: CH230-34M7.TV
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_eirng_information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 34 row: M column: 7
Seq primer: SP6
Class: BAC ends.

FEATURES
Location/Qualifiers
source 1..725
/organism="Rattus norvegicus"
/strain="BN/SSNhsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-34M7"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site_1: EcORI; Site_2: EcORI;
CHORI-230 Rat (BN/SSNhsd/MCw) BAC library produced by
Pieter de Jong"

BASE COUNT 115 a 163 c 181 g 266 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
8.00	100.00%	725	8
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	53.33%	Mismatches:	0
DB:	17	Indels:	0
		Gaps:	0

SEQ1-65to79 (1-15) x BH271552 (1-725)

OY 1 SerLeuArgSerAlaHisLeuAla 8
|||||

Db 670 AGCCTCAGGAGTGCACCACTGGCC 693

RESULT 13

LOCUS BH351398 735 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-175A4.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-175A4, DNA sequence.
ACCESSION BH351398
VERSION BH351398.1 GI:17282132
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 735)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregorjais, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P., and Fraser, C. M.
Rat BAC End Sequences from Library CHORI-230 EcORI segment
Unpublished (1999)
Other_GSSs: CH230-175A4.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/orering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 175 row: A column: 4
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..735
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone_lib="CH230-175A4"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 117 a 171 c 182 g 265 t

ORIGIN

Alignment Scores:
Pred. No.: 630 Length: 735
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 17 Gaps: 0

SEQ1-65T079 (1-15) x BH351398 (1-735)

OY 1 SerleuArgSerAlaHisLeuAla 8
|||||

Db 662 AGCCTCAGAGTGCACCTGGCC 665

RESULT 14
BI658656/c 768 bp mRNA linear EST 12-SEP-2001
LOCUS 603283613F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5328074 5',
DEFINITION mRNA sequence.
ACCESSION BI658656
VERSION BI658656
KEYWORDS BI658656.1 GI:15572892
SOURCE EST.
ORGANISM house mouse,
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 768)
TITLE NIH-MGC <http://mgc.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1832 row: 1 column: 03
High quality sequence stop: 768.

FEATURES
source
1..768
Location/Qualifiers
/organism="Mus musculus"
/strain="NMRI"
/db_xref="taxon:10090"
/clone_lib="IMAGE:5328074"
/clone_lib="NCI_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Priscilla Furch, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."

BASE COUNT 185 a 208 c 236 g 139 t

ORIGIN

Alignment Scores:
Pred. No.: 665 Length: 768
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 13 Gaps: 0

SEQ1-65T079 (1-15) x BI658656 (1-768)

OY 1 SerleuArgSerAlaHisLeuAla 8
|||||

Db 711 AGCCTGAGTGTGCACACTGGCG 688

RESULT 15
CNS03A1W 790 bp DNA linear GSS 15-MAY-2000
LOCUS 008B16 of library G from Tetraodon nigroviridis, genomic survey
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
sequence.
ACCESSION AL234653
VERSION AL234653.1 GI:7893788
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Acanthomorpha; Acanthopterygii; Teleostei; Euteleostei; Neoteleostei; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 790)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brotlier, P., Quettier, F., Saurin, W., and Weissenbach, J.
TITLE human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 790)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quettier, F., Saurin, W., Bernot, A. and Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 790)

AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000)
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
 source 1..790
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : COBG008DA08LP1-end : 77"

BASE COUNT 168 a 239 c 204 g 158 t 21 others

ORIGIN

Alignment Scores:
 Pred. No.: 688 Length: 790
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 17 Gaps: 0

SEQ1-65to79 (1-15) x CNS03A1W (1-790)

OY 2 LeuArgSerAlaHisIleuAlaGly 9
 Db 23 CTGAGGTCGCTCACCCTGGCTGCT 46

RESULT 16
 A0745347 816 bp DNA linear GSS 16-JUL-1999
 LOCUS HS_2276_AL.B11_T7C CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=2276 Col=21 Row=C, DNA sequence.
 ACCESSION A0745347
 VERSION A0745347.1 GI:5522869
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 816)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 JOURNAL 99380589
 MEDLINE
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: <http://www.husc.washington.edu>
 Plate: 2276 Row: C Column: 21
 Seq primer: 17
 Class: BAC ends
 High quality sequence stop: 816.

FEATURES
 source 1..816
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate=2276 Col=21 Row=C
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 230 a 184 c 189 g 203 t 10 others

ORIGIN

Alignment Scores:
 Pred. No.: 716 Length: 816
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 17 Gaps: 0

SEQ1-65to79 (1-15) x A0745347 (1-816)

OY 7 LeuAlaGlyInserIleLeuSer 14
 Db 600 CTGGCAGGCGCAACTATTCATCT 577

RESULT 17
 CNS03WY2 829 bp DNA linear GSS 18-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
 DEFINITION 065M23 of library G from Tetraodon nigroviridis, genomic survey sequence.
 ACCESSION AL264356.1 GI:7986049
 VERSION AL264356.1
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
 1 (bases 1 to 829)
 Roest-Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
 Unpublished
 JOURNAL 2 (bases 1 to 829)
 REFERENCE Roest-Crolius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billaud,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 3 (bases 1 to 829)
 JOURNAL Genoscope.
 REFERENCE Direct Submission
 JOURNAL Submitted (12-APR-2000)
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
 source 1..829
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="065M23"
 /clone_lib="G"
 /note="Genoscope sequence ID : COBG065AG13SP1-end : PUC-ori"

BASE COUNT 199 a 218 c 184 g 225 t 3 others

ORIGIN

Alignment Scores:
 Pred. No.: 730 Length: 829
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 17 Gaps: 0

SEQ1-65T079 (1-15) x CNS03MYZ (1-829)

OY 1 SerleuArgSerAlaHisLeuAla 8
 |||||||
 Db 583 TCTTTGAGAGTCTCTCATTTGGCC 560

RESULT 18
 BFL20183/c
 LOCUS 870 bp mRNA linear EST 24-OCT-2000
 DEFINITION 601756489F1 NCL_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3985622 5',
 mRNA sequence.
 ACCESSION BFL20183
 VERSION BFL20183.1 GI:10959223
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 870)
 NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: LLM9190 row: a column: 23
 High quality sequence stop: 651.

FEATURES
 source
 1..870
 /organism="Mus musculus"
 /strain="C57BL/6j"
 /db_xref="taxon:10090"
 /clone="IMAGE:3985622"
 /clone_lib="NCL_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site:1: SalI;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dr.
 library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 214 a 229 c 268 g 159 t
 ORIGIN

Alignment Scores:
 Pred. No.: 774 Length: 870
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 12 Gaps: 0

SEQ1-65T079 (1-15) x BFL20183 (1-870)

OY 1 SerleuArgSerAlaHisLeuAla 8
 |||||||
 Db 322 AGCCTGAGTCTCTCACACTTGGCG 299

RESULT 19
 BF575766/c
 LOCUS 884 bp mRNA linear EST 12-DEC-2000
 DEFINITION 602135396F1 NTH_MGC_81 Homo sapiens cDNA clone IMAGE:4290670 5',
 mRNA sequence.
 ACCESSION BF575766
 VERSION BF575766.1 GI:11649478
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 884)
 NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 CDNA Library Preparation: CLONETECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: LLM1133 row: o column: 23
 High quality sequence stop: 486.

FEATURES
 source
 1..884
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4290670"
 /clone_lib="NTH_MGC_81"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site:1: SfiI (ggccgctcggcc); Site:2: SfiI
 (ggccatcggcc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CACGCCATTTATGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."

BASE COUNT 225 a 163 c 267 g 229 t
 ORIGIN

Alignment Scores:
 Pred. No.: 789 Length: 884
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 12 Gaps: 0

SEQ1-65T079 (1-15) x BF575766 (1-884)

OY 1 SerleuArgSerAlaHisLeuAla 8
 |||||||
 Db 204 TCTTTAGCTCTCGCCACCTAGCA 181

RESULT 20
 CNS03N3Z
 LOCUS 967 bp DNA linear GSS 17-MAY-2000
 DEFINITION Tetradodon nigroviridis genome survey sequence T7 end of clone
 039K05 of library G from Tetradodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL251576
 VERSION AL251576.1 GI:7972568
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradodon nigroviridis.
 ORGANISM Tetradodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 967)
 Roest-Crolius H., Jallou O., Dasilva C., Bouneau L., Fisher C.,
 Bernot A., Fitzames C., Winkler P., Brotier P., Quetier F.,
 Saurin W. and Weissenbach J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 967)

AUTHORS Roest-Crollius,H., Tallon,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and Weissbach,U.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 967)

AUTHORS Genoscope.

JOURNAL Direct Submission

COMMENT Submitted (12-Apr-2000)

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source Location/Qualifiers

1..967

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone_lib="G"

/note="Genoscope sequence ID : CGG039AF03LP1-end : T7"

BASE COUNT 272 a 203 c 245 g 246 t 1 others

ORIGIN

Alignment Scores:

Pred. No.:	881	Length:	967
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	53.33%	Indels:	0
DB:	17	Gaps:	0

SEQ1-65T079 (1-15) x CNS03N32 (1-967)

QY 1 SerLeuArgSerAlaHisLeuAla 8

DB 114 TCTTTCAGAGTGCATTTGGCC 137

RESULT 21

BC019121/c 1877 bp mRNA linear HTC 07-AUG-2002

LOCUS BC019121

DEFINITION Mus musculus, clone IMAGE:5002987, mRNA.

ACCESSION BC019121

VERSION BC019121.1 GI:17403065

KEYWORDS HTC.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1877)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (07-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nih.gov

Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,

Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lalic,P., Legaspi,R., Maduro,O.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantipop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherly,K.D., Wiggins,L., Young,A., Zhang,L.H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>

Series: IRAC Plate: 39 Row: c Column: 15

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein

This clone has the following problem: no 5' EST match.

FEATURES

source Location/Qualifiers

1..1877

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="FVB/N-3"

/clone="IMAGE:5002987"

/tissue_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month old mouse, taken by biopsy."

/clone_lib="NCLCGAP_Mam2"

/lab_host="DH10B"

/note="Vector: PCMV-SPOrt6"

BASE COUNT 473 a 504 c 525 g 375 t

ORIGIN

Alignment Scores:

Pred. No.:	1,98+03	Length:	1877
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	53.33%	Indels:	0
DB:	11	Gaps:	0

SEQ1-65T079 (1-15) x BC019121 (1-1877)

QY 1 SerLeuArgSerAlaHisLeuAla 8

DB 353 AGCCTGAGTCTGCACACTTGCC 330

Search completed: January 28, 2003, 13:16:52

Job time : 1272.67 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:38:10 : Search time 54 Seconds

(without alignments)
85.188 Million cell updates/sec

Title: SEQ1-65T079

Perfect score: 15
Sequence: 1 SLRSAHLAQSTILSG 15

Scoring table:

	OLIGO
Xgapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 441362 seqs, 15338381 residues

Word size: 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=framer_p2n.model -DEV=rlh
-Q=/cgn2_1/USPTO.spool/BORIN682/runat_23012003_130141_7891/app.query.fasta_1.597
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=olip2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCCALIGN=200 -THR.SCORE=quality -THR.MIN=8 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=200000000
-USRR=BORIN682_6CGN_1_1_25_6runat_23012003_130141_7891 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_XMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	10	66.7	711	3	US-08-823-120-5

ALIGNMENTS

RESULT 1
US-08-823-120-5
Sequence 5, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino

APPLICANT: Pizze, Mariagrazia
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
TITLE OF INVENTION: Cholera Toxin and of the Toxin Lt. Their Preparation and
TITLE OF INVENTION: Their Use for the Preparation of Vaccines
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 711 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..711
US-08-823-120-5

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0	0.0182	10.00	100.00%	66.67%	711	10	0	0	0	0

SEQ1-65T079 (1-15) x US-08-823-120-5 (1-711)

QY 1 SerLeuArgSerAlaHisLeuAlaGlyGln 10
|||||
Db 190 AGTTGAGAGTCTCTACTTACGACGACAG 219

Search completed: January 28, 2003, 13:19:48
Job time : 54 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:39:05 ; Search time 63 seconds
(without alignments)
106.968 million cell updates/sec

Title: SEQ1-65TO79
Perfect score: 15
Sequence: 1 SLRSAMLAGQSILSG 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 396772 seqs, 224632407 residues

Word size: 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+_p2n_model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/BORIN682/rnat_23012003_130142_7908/app-query.fasta.1.597
-DB=published.Applications_NA -QFWT=fastap -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=8
-ALIGN=50 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=BORIN682 @cgn_1.1.33 @rnat_23012003_130142_7908
-NCPU=6 -ICPU=3 -NO_XLIPX -NO_WMAP -LARGEDUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	723	9 US-09-950-335A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-950-335A-5
; Sequence 5, Application US/09950335A
; Publication No. US20020193330A1
; GENERAL INFORMATION:
; APPLICANT: HONE, DAVID M.
; TITLE OF INVENTION: GENETICALLY ENGINEERED CO-EXPRESSION DNA VACCINES, CONSTRUCTIO
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 4115-128
; CURRENT APPLICATION NUMBER: US/09/950,335A
; CURRENT FILING DATE: 2001-09-10
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-950-335A-5

Alignment Scores:
Pred. No.: 1.4e-07
Score: 15.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 9
Gaps: 0

SEQ1-65TO79 (1-15) x US-09-950-335A-5 (1-723)

Qy 1 SerLeuArgSerAlaHisLeuAlaGlyGlnSerIleLeuSergly 15
|||||
Db 193 AGTTTGAGAACTGCTCACTTAGCAGACAGCTATATATATCAGCA 237

Search completed: January 28, 2003, 13:23:11
Job time : 63 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 28, 2003, 10:33:50 ; Search time 1527.67 seconds
(without alignments)
285.757 Million cell updates/sec

Title: SEQ3-65T079
Perfect score: 15
Sequence: 1 STFEQVPPNKEFKGV 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Word size: 8
Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 1000 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPRO/BORIN82/runat_23012003_130141_7873/app.query.fasta.1.597
-DB=geneml -QWRT=fastap -SUFF=olip2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000
-LOCALIGN=200 -THR.SCORE=quality -THR.MIN=8 -ALIGN=50 -MODE=LOCAL -OUTFMT=pro
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=BORIN82 @CGN_1.1.2425 @runat_23012003_130141_7873 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	60.0	1200	1 ECOLTIIA	M17894 E.coli heat
2	8	53.3	229289	10 AL603842	AL603842 Mouse DNA

ALIGNMENTS

RESULT 1
ECOLTIIA
LOCUS
DEFINITION E.coli heat-labile enterotoxin type IIA (LT-IIa) A and B genes,
complete cds.
ACCESSION M17894
VERSION M17894.1 GI:146671
KEYWORDS enterotoxin; heat-labile enterotoxin.
SOURCE E.coli (strain SA53) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1200)
Pickett,C.L., Weinstein,D.L. and Holmes,R.K.
TITLE Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
JOURNAL Operon fusions, nucleotide sequence, and hybridization studies
MEDLINE J. Bacteriol. 169 (11), 5180-5187 (1987)
PUBMED 88032841
FEATURES
source
1..1200 location/Qualifiers
/organism="Escherichia coli"
/db_xref="taxon:562"
60..839
/note="heat labile enterotoxin type IIA A"
/codon_start=1
/transl_table=11
/protein_id="AA024093.1"
/db_xref="GI:146672"
/translation="MIKIVLLFVPIFSVSANDFFRADSRTPDEIRRAGCILPRGQO
EAYERGPININLYEHARGTVAGRYNDGVYSTVTLRQAHILICONTLGSINEYIY
VAPAPNLFVNGVILGRVSPYSENEFPAALGIPISOTIGWYRVSEFGLIEGSMORND
VRGDLFRGILTVAPNEDGVOAGFSPNFRAMKREMPSTAPDQCVNKNKEFKGVCISA
TNVLSKIDLMNFKILKRRLLATFPMSDDFLGVHGERDEL"
829..1200
/note="heat labile enterotoxin type IIA B"
/codon_start=1
/transl_table=11
/protein_id="AA024094.1"
/db_xref="GI:146673"
/translation="WSSKRIIGAFVLMTGILSGOVYACVSEHFRNICOTTADIVAGV
OLKRYIADVNTNRGIIYVSNVTGCGWVYIPGGRDYPDNFLSGIRKTAAAILSDTKVN
LCARSSSPNHIWAMELDRES"

BASE COUNT 373 a 175 c 294 g 358 t

ORIGIN

Alignment Scores:

Pred. No.:	0.198	Length:	1200
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	60.00%	Indels:	0
DB:	1	Gaps:	0

SEQ3-65TO79 (1-15) x ECOLIITA (1-1200)

QY 6 ValProAsnAnlysgluPhelysGly 14

Db 669 GTGCCGAATATAAGATTAAAGGA 695

RESULT 2

AL603842

LOCUS AL603842 229289 bp DNA linear ROD 17-NOV-2001
DEFINITION Mouse DNA sequence from clone RP23-100P23 on chromosome 11,
complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (17-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Oct 30, 2001 this sequence version replaced gi:16304764. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Mp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-100P23 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

This sequence is the entire insert of clone RP23-100P23.

FEATURES

Source

1..229289

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="11"

/clone="RP23-100P23"

/clone_lib="RPI-23"

59782..59788

/note="Sequence from uni-directional dGTP big dye terminator reads only."

125096..125397

/note="Single clone region. Sequence from clone PCR only."

170770..170774

/note="1327 Bases of IS2 transposon (V00610) removed here. This sequence represents the duplicated flanking sequence

misc_feature 176557..176641

/note="Sequence from uni-directional dGTP big dye terminator reads only."

BASE COUNT 64835 a 50446 c 50226 g 63782 t

ORIGIN

Alignment Scores:

Pred. No.:	136	Length:	229289
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	53.33%	Indels:	0
DB:	10	Gaps:	0

SEQ3-65TO79 (1-15) x AL603842 (1-229289)

QY 6 ValProAsnAnlysgluPhelys I3

Db 11390 GTACCAACATATAAGAGTTTAAAG 11413

Search completed: January 28, 2003, 12:15:31
Job time : 1536.67 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:32:26 : Search time 344.333 Seconds
(without alignments)
98.103 Million cell updates/sec

Title: SEQ3-65to79
Perfect score: 15
Sequence: 1 STEEQVNNKEFGKV 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Word size: 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xih
-Q=cg2_1/USPTO.spool/BORIN682/runat_23012003_130140.7866/app.query.fasta_1.597
-DB=N-Geneseq_101002 -QFMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdl
-LIST=1000 -DOCCALIGN=200 -THR.SCORE=quality -THR_MIN=8 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=200000000
-USFR=BORIN682_66GN_1_1.187_@runat_23012003_130140.7866 -NCPU=6 -ICPU=3
-NO_XLFX -NO_JMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEDUT=120
-MARN_TIMEDUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
No matches found					

Search completed: January 28, 2003, 10:56:27
Job time : 344.333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:36:45 ; Search time 1270.67 Seconds
(without alignments)
191.185 Million cell updates/sec

Title: SEQ3-65to79
Perfect score: 15
Sequence: 1 STEEQVNNKEKGV 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODE=frame+ p2n model -DEV=ylh
-Q/cgq2.1/USFTO.spool/BORN62/unat_23012003.130141.7880/app.query.fasta.1.597
-DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -DOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cgi -LIST=1000
-DOCALLIGN=200 -THR.SCORE=quality -THR_MIN=8 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=BORN62 @CGN.1.1.1349 @runat.23012003.130141.7880 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAMP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -TTHREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8	53.3	683	10	BB016585	BB016585 BB016585

ALIGNMENTS

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
BB016585	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
LOCUS	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
DEFINITION	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
ACCESSION	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
VERSION	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
KEYWORDS	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
SOURCE	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
ORGANISM	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
REFERENCE	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	
AUTHORS	BB016585	BB016585	BB016585	BB016585.2	GI:16257466	house mouse.	EST.	

BB016585 683 bp mRNA linear EST 18-OCT-2001
BB016585 RIKEN full-length enriched, adult male testis (DH10B) Mus
musculus cDNA clone 4930563A03 3', mRNA sequence.
BB016585
BB016585.2 GI:16257466
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 683)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jun 2, 2000 this sequence version replaced gi:8187724.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
waga,K., Fujikake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,K., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1. 683
source

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930563A03"
/clone_lib="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTATTAATATATCCCCCCCCCCC 3']. cDNA
was cloned into the xhoI and BamHI sites. Vector: a
modified pluscript KS(+) after bulk excision from Lambda
FLC 1. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT 223 a 124 c 141 g 195 t
ORIGIN

Alignment Scores:
Pred. No.: 11.2 Length: 683
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 10 Gaps: 0

SEQ3-65to79 (1-15) x BB016585 (1-683)

QY 6 ValProAsnAsnLysGluPheLys 13
|||||
DB 660 GTACCAACAATTAAGAGTTTAAG 683

Search completed: January 28, 2003, 13:16:52
Job time : 1270.67 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:38:10 ; Search time 54 Seconds
(without alignments)
85.188 Million cell updates/sec

Title: SEQ3-65TO79
Perfect score: 15
Sequence: 1 STEEQVPMNKEFGV 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 1000 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=x1h
-O=/cgn2_1/USFTO.spool/BORIN682/runat_23012003_130141_7891/app_query.fasta_1.597
-DB=Issued_Patents_NA_QFMT=fastap -SUFFIX=olip2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCCALIGN=200 -THR.SCORE=quality -THR.MIN=8 -ALIGN=50 -MODE=LOCAL
-OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=BORIN682 @CGN_1.1.25 @runat_23012003_130141_7891 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
No matches found					

Search completed: January 28, 2003, 13:19:48
Job time : 54 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

Search completed: January 28, 2003, 13:23:11
Job time : 63 secs

OW protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:39:05 : Search time 63 seconds
(without alignments)
106.968 Million cell updates/sec

Title: SEQ3-65TO79
Perfect score: 15
Sequence: 1 STFEQVPNNKEFGV 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 396772 seqs, 224632407 residues

Word size: 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USFTO/spool/BORIN682/unat_23012003_130142_7908/app_query.fasta_1.597
-DB=published_Applications_NA -QMT=fastcap -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human4.0.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=8
-ALIGN=50 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=BORIN682@cgn_1_1_33@unat_23012003_130142_7908
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description
No.	Score				
No matches found					

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:33:50 ; Search time 1527.67 Seconds
(without alignments)
285.757 Million cell updates/sec

Title: SEQ4-65to79
Perfect score: 15
Sequence: 1 REFNSLPNKASSDT 15

Scoring table:
Oligo
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 8
Total number of hits satisfying chosen parameters: 13

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xih
-O/cg92_1/USPTO.spool/BORIN682/runat.23012003.130141.7873/app_query.fasta.1.597
-DB=GenEmbl -GFWT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -DOOPCL=0 -DOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=olip2n -TRANS=human40.cdi -LIST=1000
-DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=8 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER-BORIN682 @CGN.1.1.2425 @runat.23012003.130141.7873
-NORM_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-NORM.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rod:*
36: em.htg_mem:*
37: em.htg_vrt:*
38: em.sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	66.7	1262	1 ECOTOXHL	M28523 Escherichia
2	8	53.3	635	10 AF204796S8	AF204803 Mus muscu
3	8	53.3	71402	2 AC123728	AC123728 Mus muscu
4	8	53.3	101445	9 AL133411	AL133411 Human DNA
5	8	53.3	114928	2 AC125479	AC125479 Medicago
6	8	53.3	142959	2 AC025453	AC025453 Homo sapi
7	8	53.3	145882	2 AL159131	AL159131 Homo sapi
8	8	53.3	152617	2 AC113881	AC113881 Rattus no
9	8	53.3	154455	2 AC021844	AC021844 Homo sapi
10	8	53.3	155394	2 AC092328	AC092328 Homo sapi
11	8	53.3	176123	9 AL390718	AL390718 Human DNA
12	8	53.3	184635	9 AC025445	AC025445 Homo sapi
13	8	53.3	192263	10 AC068609	AC068609 Mus muscu

ALIGNMENTS

RESULT 1
ECOTOXHL
LOCUS 1262 bp DNA linear BCT 18-NOV-1994
DEFINITION Escherichia coli (strain 41) heat-labile enterotoxin type IIB
(LT-IIB) A and B chain genes, complete cds.

ACCESSION M28523.1 GI:576584
VERSION M28523
KEYWORDS LT-IIB gene; enterotoxin type IIB.
SOURCE Escherichia coli (strain 41) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 1262)
Pickett,C.L., Twiddy,E.M., Coker,C. and Holmes,R.K.
TITLE Cloning, nucleotide sequence, and hybridization studies of the type
IIB heat-labile enterotoxin gene of Escherichia coli

JOURNAL J. Bacteriol. 171 (9), 4945-4952 (1989)

MEDLINE 89359131

PUBMED 2670900

COMMENT On Nov 28, 1994 this sequence version replaced g1:341953.
FEATURES
Location/Qualifiers
1..1262
/organism="Escherichia coli"
/strain="41"
/db_xref="taxon:562"
/clone="pcp4185"
1..1262
/gene="LT-IIB"
1..12
/gene="LT-IIB"
26..31
/gene="LT-IIB"
113..904

gene

-35_signal

-10_signal

CDS

```

/gene="LT-IIb"
/note="A chain of heat-labile enterotoxin type IIB"
/codon_start=1
/transl_table=11
/product="enterotoxin"
/protein_id="AA53285.1"
/db_xref="GI:576585"
/translation="MAKYISFISLFLISPLFYANDYERADSRPEDEVRSGLIPRG
ODEAYERTPINILYDHARGATGENTNGYVSTTTLRQAHLLGQNMGGNEY
IYVVAAPNLEPDVNGVLGRSPYSENEYALGGIPLSOTIGWRYVSGALTEGMRN
RDYRDLFRGISAAPNEDGYRIAGPDGFPAMVEYEMREPAFNSCLPNNKSSDTTCA
SLTNKLSQHDIADEFKRYIKRFTLMTLLSLINNDGFFSNCKDEL"
113..172
/gene="LT-IIb"
/mat_peptide 173..901
/gene="LT-IIb"
/product="enterotoxin"
/note="A chain of heat-labile enterotoxin type IIB"
894..1262
/gene="LT-IIb"
/note="B chain of heat-labile enterotoxin type IIB"
/codon_start=1
/transl_table=11
/product="enterotoxin"
/protein_id="AA53286.1"
/db_xref="GI:576586"
/translation="MSFKIIRKAFYIMALYVOAHAGASQEFKQNCNRTTASLVEG
ELTKYIDNNNDGMYVSSGGVWYRSRAKDPDNNVTMEMRIRIAAAVALSGMRV
MCASPASSPNYIMALELFAE"
894..962
/gene="LT-IIb"
/mat_peptide 963..1259
/gene="LT-IIb"
/product="enterotoxin"
/note="B chain of heat-labile enterotoxin type IIB"
BASE COUNT 369 a 195 c 285 g 413 t
ORIGIN

Alignment Scores:
Pred. No.: 0.00419 Length: 1262
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 1 Gaps: 0

SEQ4-65T079 (1-15) x ECOT0XHL (1-1262)

QY 6 LeuProAsnAsnLysAlaSerSerAspThr 15
Db 728 CTCACCAATATATTAAGCTTCAGTGATACT 757

RESULT 2
AF20479658/c AF20479658 635 bp DNA linear ROD 30-APR-2001
LOCUS Mus musculus 14-alpha-demethylase (CYP51) gene, exon 9.
DEFINITION AF204803
ACCESSION AF204803.1 GI:8347234
VERSION
KEYWORDS
SEGMENT
SOURCE 8 of 9
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 635)
AUTHORS Debeljak,N., Horvat,S., Komel,R. and Rozman,D.
TITLE Molecular cloning and partial characterisation of the mouse Cyp51
CDNA
JOURNAL Pfugers Arch. 439 (3 Suppl.), R7-R8 (2000)
MEDLINE 20116824
PUBMED 10653123
2 (bases 1 to 635)
REFERENCE Debeljak,N., Horvat,S., Vouk,K., Lee,M. and Rozman,D.
AUTHORS

```

```

TITLE
Characterization of the mouse lanosterol 14alpha-demethylase
(CYP51), a new member of the evolutionarily most conserved
cytochrome P450 family
JOURNAL Arch. Biochem. Biophys. 379 (1), 37-45 (2000)
MEDLINE 20322886
PUBMED 10864439
REFERENCE 3 (bases 1 to 635)
AUTHORS Debeljak,N., Vouk,K., Gazvoda,B. and Horvat,S.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-1999) Medical Center for Molecular Biology,
Institute of Biochemistry, Medical Faculty, University of
Ljubljana, Vrazov trg 2, Ljubljana 1000, Slovenia
FEATURES
source
1..635
/organism="Mus musculus"
/strain="129/Sv"
/db_xref="taxon:10090"
/chromosome="5"
/map="A2"
<1..345
/number=8
/gene="CYP51"
/exon 346..514
/number=9
/gene="CYP51"
/intron 515..>635
/number=9
/gene="CYP51"
BASE COUNT 170 a 111 c 144 g 210 t
ORIGIN

Alignment Scores:
Pred. No.: 0.76 Length: 635
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 10 Gaps: 0

SEQ4-65T079 (1-15) x AF20479658 (1-635)

QY 5 SerLeuProAsnAsnLysAlaSer 12
Db 305 TCACACCAACAATTAAGCATCA 282

RESULT 3
AC123728/c AC123728 71402 bp DNA linear HMG 15-JUN-2002
LOCUS Mus musculus clone RP23-453L19, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC123728
ACCESSION AC123728.2 GI:21427923
VERSION
KEYWORDS HMG; HMGs_PHASED.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 71402)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Unpublished
JOURNAL 2 (bases 1 to 71402)
REFERENCE 2 (bases 1 to 71402)
AUTHORS Anderson,S., Batta,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhvalter,B., Brown,A., Camarita,J., Campopiano,A., Chang,I.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glend,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamazares,R., Landers,T., Lenockzy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,

```


TITLE
JOURNAL
REFERENCE
AUTHORS

Meneus, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stenge-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 71402)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Bara, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Bouhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
Galegov, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lakoque, K.,
Lameres, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,
Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stenge-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 15, 2002 this sequence version replaced gi:21307290.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996,1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L26488

Center clone name: 453_L_19

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 728 827: contig of 727 bp in length
* 828 1556: gap of 100 bp
* 1557 1656: contig of 729 bp in length
* 1657 2387: gap of 100 bp
* 2388 2487: contig of 731 bp in length
* 2488 3228: gap of 100 bp
* 3229 3328: contig of 741 bp in length
* 3329 4055: gap of 100 bp
* 4056 4155: contig of 727 bp in length
* 4156 4889: gap of 100 bp

4890 4989: contig of 734 bp in length
4990 5722: gap of 100 bp
5723 5822: contig of 733 bp in length
5823 6545: gap of 100 bp
6546 6645: contig of 723 bp in length
6646 7374: gap of 100 bp
7375 7474: contig of 729 bp in length
7475 8156: gap of 100 bp
8157 8296: contig of 722 bp in length
8297 9026: gap of 100 bp
9027 9126: contig of 730 bp in length
9127 9841: gap of 100 bp
9842 9941: contig of 715 bp in length
9942 10642: gap of 100 bp
10643 10742: contig of 701 bp in length
10743 11458: gap of 100 bp
11459 11558: contig of 716 bp in length
11559 12248: gap of 100 bp
12249 12348: contig of 690 bp in length
12349 13078: gap of 100 bp
13079 13178: contig of 730 bp in length
13179 13889: gap of 100 bp
13890 13989: contig of 711 bp in length
13990 14704: gap of 100 bp
14704 14804: contig of 715 bp in length
14805 15525: gap of 100 bp
15526 15625: contig of 721 bp in length
15626 16329: gap of 100 bp
16330 16429: contig of 704 bp in length
16430 17139: gap of 100 bp
17140 17239: contig of 710 bp in length
17240 17977: gap of 100 bp
17977 18077: contig of 738 bp in length
18078 18077: gap of 100 bp
18078 18800: contig of 723 bp in length
18801 18900: gap of 100 bp
18901 19633: contig of 733 bp in length
19634 19733: gap of 100 bp
19734 20470: contig of 737 bp in length
20471 20570: gap of 100 bp
20571 21304: contig of 734 bp in length
21305 21404: gap of 100 bp
21405 22155: contig of 751 bp in length
22156 22255: gap of 100 bp
22256 22992: contig of 737 bp in length
22993 23092: gap of 100 bp
23093 23815: contig of 723 bp in length
23816 23915: gap of 100 bp
23916 24648: contig of 733 bp in length
24649 24748: gap of 100 bp
24749 25467: contig of 719 bp in length
25468 25567: gap of 100 bp
25568 26291: contig of 724 bp in length
26292 26391: gap of 100 bp
26392 27110: contig of 719 bp in length
27111 27210: gap of 100 bp
27211 27948: contig of 738 bp in length
27949 28048: gap of 100 bp
28049 28778: contig of 730 bp in length
28779 28878: gap of 100 bp
28879 29601: contig of 723 bp in length
29602 29701: gap of 100 bp
29702 30450: contig of 749 bp in length
30451 30550: gap of 100 bp
30551 31267: contig of 717 bp in length
31268 31367: gap of 100 bp
31368 32081: contig of 714 bp in length
32082 32181: gap of 100 bp
32182 32910: contig of 729 bp in length
32911 33010: gap of 100 bp
33011 33730: contig of 720 bp in length
33731 33830: gap of 100 bp
33831 34551: contig of 721 bp in length

* 34552 34651: gap of 100 bp
 * 34652 35372: contig of 721 bp in length
 * 35373 35472: gap of 100 bp
 * 35473 36191: contig of 719 bp in length
 * 36192 36291: gap of 100 bp
 * 36292 37018: contig of 727 bp in length
 * 37019 37118: gap of 100 bp
 * 37119 37849: contig of 731 bp in length
 * 37850 37949: gap of 100 bp
 * 37950 38691: contig of 742 bp in length
 * 38692 38791: gap of 100 bp
 * 38792 39511: contig of 720 bp in length
 * 39512 39611: gap of 100 bp
 * 39612 40329: contig of 718 bp in length
 * 40330 40429: gap of 100 bp
 * 40430 41168: contig of 739 bp in length
 * 41169 41268: gap of 100 bp
 * 41269 41999: contig of 731 bp in length
 * 42000 42099: gap of 100 bp
 * 42100 42834: contig of 735 bp in length
 * 42835 42934: gap of 100 bp
 * 42935 43655: contig of 721 bp in length
 * 43656 43755: gap of 100 bp
 * 43756 44479: contig of 724 bp in length
 * 44480 44579: gap of 100 bp
 * 44580 45326: contig of 747 bp in length
 * 45327 45426: gap of 100 bp
 * 45427 46155: contig of 729 bp in length

Alignment Scores:

Pred. No.: 44.1 Length: 71402
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 2 Gaps: 0

SEQ4-65TO79 (1-15) x AC123728 (1-71402)

Qy 4 AsnSerLeuProAsnAsnLys 11

Db 2669 AATAGTCTCCCAACATAAGGCC 2646

RESULT 4
 AL133411/c 101445 bp DNA linear PRI 17-MAR-2002
 LOCUS Human DNA sequence from clone RP11-57P14 on chromosome 9p21.1-21.3,
 DEFINITION complete sequence.

ACCESSION AL133411
 VERSION AL133411.9 GI:15572876
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE 1 (bases 1 to 101445)
 AUTHORS Corby, N.
 JOURNAL Direct Submission
 COMMENT Submitted (17-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
 On Mar 21, 2002 this sequence version replaced gi:10931831.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/chr9>
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP11-57P14 is from
 the library RP11-11.1 constructed by the group of Pieter de Jong.
 For further details see
<http://www.chori.org/dacpac/home.htm>
 VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-57P14 it may be shorter because we sequenced overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-298B2 is at 9946 in this sequence.
 The true right end of clone RP11-179D22 is at 2000 in this
 sequence.

FEATURES

Source

1. 101445
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="p21.1-21.3"
 /clone="RP11-57P14"
 /clone_id="RP11-11.1"
 BASE COUNT 33258 a 20082 c 19978 g 28127 t
 ORIGIN

Alignment Scores:
 Pred. No.: 59.7 Length: 101445
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 9 Gaps: 0

SEQ4-65TO79 (1-15) x AL133411 (1-101445)

Qy 3 PheAsnSerLeuProAsnAsnLys 10

Db 95205 TTCACATCCCTACACCAACACAA 95182

RESULT 5
 AC125479 114928 bp DNA linear HTG 20-AUG-2002
 LOCUS Medicago truncatula clone mth1-7f11, WORKING DRAFT SEQUENCE, 7
 DEFINITION ordered pieces.

ACCESSION AC125479
 VERSION AC125479.5 GI:22297354
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.

TITLE 1 (bases 1 to 114928)
 AUTHORS Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
 Cook, D., Kim, D. and Roe, B.A.
 JOURNAL Medicago truncatula BAC Clone mth1-7f11
 REFERENCE 2 (bases 1 to 114928)
 AUTHORS Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
 Cook, D., Kim, D. and Roe, B.A.
 JOURNAL Direct Submission

COMMENT Submitted (27-JUN-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 REFERENCE 3 (bases 1 to 114928)

AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2002) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 19, 2002 this sequence version replaced g1:22138503.

Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 3787: contig of 3787 bp in length
* 3788 3687: gap of unknown length
* 3888 13161: contig of 9274 bp in length
* 13162 13261: gap of unknown length
* 13262 23764: contig of 10503 bp in length
* 23765 23864: gap of unknown length
* 23865 37301: contig of 13437 bp in length
* 37302 37401: gap of unknown length
* 37402 51846: contig of 14445 bp in length
* 51847 51946: gap of unknown length
* 51947 73672: contig of 21726 bp in length
* 73673 73773 114928: gap of unknown length
* 73773 114928: contig of 41156 bp in length.
Location/Qualifiers
1. 114928
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="mhl-7f11"
/clone.lib="Medicago truncatula BAC library H1"
BASE COUNT 38474 a 18700 c 18036 g 39117 t 601 others
ORIGIN
Alignment Scores:
Pred. No.: 66.4 Length: 114928
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0
SEQ4-65to79 (1-15) x AC125479 (1-114928)
QY 2 GIUpheAnSerLeuProAnAsn 9
Db 9568 GAATCAACTCTTACCCACAAT 9591
RESULT 6
AC025453/c 142959 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2150A8, WORKING DRAFT SEQUENCE,
DEFINITION 8 ordered pieces.
AC025453
AC025453 5 GI:13677022
VERSION AC025453.5
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE DOE Joint Genome Institute.
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
Unpublished

REFERENCE 2 (bases 1 to 142959)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 19, 2001 this sequence version replaced g1:9256482.

Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 685157
Center clone name: CITB-HL_2150A8

Summary Statistics
Consensus quality: 138699 bases at least Q40
Consensus quality: 141252 bases at least Q30
Consensus quality: 141900 bases at least Q20
Estimated insert size: 155560; agarose-1p estimation
Estimated insert size: 142259; sum-of-contigs estimation
Quality coverage: 5.9 in Q20 bases; agarose-1p estimation
Quality coverage: 6.45 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 5661: contig of 5661 bp in length
* 5662 5761: gap of unknown length
* 5762 23715: contig of 17954 bp in length
* 23716 23815: gap of unknown length
* 23816 83941: contig of 60126 bp in length
* 83942 84041: gap of unknown length
* 84042 97947: contig of 13906 bp in length
* 97948 98047: gap of unknown length
* 98048 104249: contig of 6202 bp in length
* 104250 104349: gap of unknown length
* 104350 106891: contig of 2542 bp in length
* 106892 106991: gap of unknown length
* 106992 135029: contig of 28038 bp in length
* 135030 135129: gap of unknown length
* 135130 142959: contig of 7830 bp in length.
Location/Qualifiers
1. 142959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2150A8"
/clone.lib="Caitrech human BAC library D"
BASE COUNT 47280 a 26024 c 25587 g 43364 t 704 others
ORIGIN
Alignment Scores:
Pred. No.: 80.2 Length: 142959
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.33% Indels: 0
DB: 2 Gaps: 0
SEQ4-65to79 (1-15) x AC025453 (1-142959)
QY 3 PheAnSerLeuProAnAsnLys 10
Db 139734 TTTAACCTTTGGCTAATACAAAG 139711
RESULT 7
AL139131

```

LOCUS      AL139131      145882 bp      DNA      linear      HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-172P10, *** SEQUENCING IN
PROGRESS ***, 11 unordered pieces.
ACCESSION  AL139131
VERSION    AL139131.5  GI:9796332
KEYWORDS   HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 145882)
AUTHORS   Plumb,B.
TITLE     Direct Submission
JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            Requests: clonerequests@sanger.ac.uk
            On Aug 11, 2000 this sequence version replaced gi:9212172.
COMMENT   ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            ----- Project Information
            Center project name: ba172p10
            ----- Summary Statistics
            Sequencing program: XGAP4; version 4.5
            Sequencing vector: plasmid; 108752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 140505 bases at least Q40
            Consensus quality: 142252 bases at least Q30
            Insert size: 144882; sum-of-contigs
            Insert size: 156750; 4.8% error; agarose-fp
            Quality coverage: 3.78x in Q20 bases; sum-of-contigs Quality
            coverage: 3.58x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            -----
            1 28658: contig of 28658 bp in length
            * 28659 28758: gap of 100 bp
            * 28759 33908: contig of 5150 bp in length
            * 33909 34008: gap of 100 bp
            * 34009 37706: contig of 3698 bp in length
            * 37707 37806: gap of 100 bp
            * 37807 50897: contig of 13091 bp in length
            * 50898 50997: gap of 100 bp
            * 50998 56577: contig of 5580 bp in length
            * 56578 56677: gap of 100 bp
            * 56678 66988: contig of 10311 bp in length
            * 66989 67088: gap of 100 bp
            * 67089 69189: contig of 2101 bp in length
            * 69190 69289: gap of 100 bp
            * 69290 103697: contig of 34408 bp in length
            * 103698 103797: gap of 100 bp
            * 103798 116670: contig of 12873 bp in length
            * 116671 116770: gap of 100 bp
            * 116771 131840: contig of 15070 bp in length
            * 131841 131940: gap of 100 bp
            * 131941 145882: contig of 13942 bp in length.
FEATURES
Source
1. 145882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-172P10"
/clone_id="RPCT-11.1"
1. 28658
misc_feature

```

```

/note="assembly-fragment:01464
clone_end:T7
vector_side:left"
misc_feature
28759..33908
/note="assembly-fragment:00982
fragment_chain:1"
34009..37706
/note="assembly-fragment:00350
fragment_chain:1"
37807..50897
/note="assembly-fragment:00363"
50998..56577
/note="assembly-fragment:00444"
56678..66988
/note="assembly-fragment:00610"
67089..69189
/note="assembly-fragment:00808"
69290..103697
/note="assembly-fragment:01135..0"
103798..116670
/note="assembly-fragment:01400"
116771..131840
/note="assembly-fragment:01402"
131941..145882
/note="assembly-fragment:00106
clone_end:SP6
vector_side:right"
BASE COUNT  37225 a 34072 c 34219 g 39350 t 1016 others
ORIGIN
Alignment Scores:
Pred. No.:      81.6      Length:      145882
Score:          8.00      Matches:      8
Percent Similarity: 100.00%  Mismatches: 0
Best Local Similarity: 100.00%  Indels:      0
Query Match:     53.33%      Gaps:        0
DB:              2
SEQ4-65T079 (1-15) x AL139131 (1-145882)
OY      5 SerleupProAmsnlysa1aser 12
Db      27096 TCCTCCCAATACAAAGCATCA 27119
RESULT 8
AC113881      152617 bp      DNA      linear      HTG 17-JUL-2002
LOCUS      AC113881
DEFINITION Rattus norvegicus clone CH230-33965, *** SEQUENCING IN PROGRESS
***, 47 unordered pieces.
ACCESSION  AC113881
VERSION    AC113881.3  GI:21745684
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 152617)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbarella,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

```

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsl,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kurshl,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Liew,C., Liu,J., Liu,W., Louised,H.,
Lozdo,R.D., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwu,S., Ogih,M., Okunnu,G.,
Oreagunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Plimus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,T.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanli,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Unpublished
2 (bases 1 to 152617)
Worley,K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 152617)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gl:19526054.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GS2S
Center clone name: CH230-339G5
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 118822 bases at least Q40
Consensus quality: 122734 bases at least Q30
Consensus quality: 125850 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_Graft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1078: contig of 1078 bp in length
* 1079 1178: gap of unknown length
* 1179 2244: contig of 1066 bp in length
* 2245 2344: gap of unknown length
* 2345 4170: contig of 1826 bp in length
* 4171 4270: gap of unknown length
* 4271 5512: contig of 1242 bp in length
5513 5612: gap of unknown length
5613 6797: contig of 1185 bp in length
6798 6897: gap of unknown length
6898 8080: contig of 1183 bp in length
8081 8180: gap of unknown length
8181 9363: contig of 1183 bp in length
9364 9463: gap of unknown length
9464 10642: contig of 1179 bp in length
10643 10742: gap of unknown length
10743 11923: contig of 1181 bp in length
11924 12023: gap of unknown length
12024 13180: contig of 1157 bp in length
13181 13280: gap of unknown length
13281 14475: contig of 1195 bp in length
14476 14575: gap of unknown length
14576 16712: contig of 2137 bp in length
16713 16812: gap of unknown length
16813 17994: contig of 1182 bp in length
17995 18094: gap of unknown length
18095 19970: contig of 1876 bp in length
19971 20070: gap of unknown length
20071 21228: contig of 1158 bp in length
21229 21328: gap of unknown length
21329 23273: contig of 1945 bp in length
23274 25654: gap of unknown length
25655 25754: gap of unknown length
25755 26848: contig of 1094 bp in length
26849 26948: gap of unknown length
26949 28666: contig of 1718 bp in length
28667 28766: gap of unknown length
28767 30806: contig of 2040 bp in length
30807 30906: gap of unknown length
30907 32637: contig of 1731 bp in length
32638 32737: gap of unknown length
32737 32737: gap of unknown length
32738 35237: contig of 2500 bp in length
35238 35337: gap of unknown length
35338 37925: contig of 2588 bp in length
37926 38025: gap of unknown length
38026 39373: contig of 1348 bp in length
39374 39473: gap of unknown length
39473 42277: contig of 2804 bp in length
42278 42377: gap of unknown length
42378 44081: contig of 1704 bp in length
44082 44181: gap of unknown length
44182 46167: contig of 1968 bp in length
46168 46267: gap of unknown length
46268 49359: contig of 3092 bp in length
49360 49459: gap of unknown length
49460 51861: contig of 2402 bp in length
51862 51961: gap of unknown length
51961 55032: contig of 3071 bp in length
55033 55132: gap of unknown length
55133 57379: contig of 2247 bp in length
57380 57479: gap of unknown length
57480 60974: contig of 3485 bp in length
60975 61074: gap of unknown length
61074 65403: contig of 4329 bp in length
65404 65503: gap of unknown length
65504 68824: contig of 3321 bp in length
68825 68924: gap of unknown length
68925 74417: contig of 5493 bp in length
74418 74517: gap of unknown length
74518 77433: contig of 2916 bp in length
77434 77533: gap of unknown length
77534 80827: contig of 3194 bp in length
80828 86128: gap of unknown length
86129 86228: gap of unknown length
86229 90700: contig of 4472 bp in length
90701 95758: gap of unknown length
95759 95858: contig of 4958 bp in length
95859 95759: gap of unknown length

```

* 95859 99072: contig of 3214 bp in length
* 99073 99172: gap of unknown length
* 99173 103575: contig of 4403 bp in length
* 103576 103675: gap of unknown length
* 103676 112734: contig of 9059 bp in length
* 112735 112834: gap of unknown length
* 112835 119742: contig of 6908 bp in length
* 119743 119842: gap of unknown length
* 119843 129603: contig of 9761 bp in length
* 129604 129703: gap of unknown length
* 129704 138726: contig of 9023 bp in length
* 138727 138826: gap of unknown length
* 138827 152617: contig of 13791 bp in length.

```

FEATURES
source
1. 152617
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-33965"

BASE COUNT 38935 a 31543 c 31335 g 40206 t 10598 others

ORIGIN

Alignment Scores:

Pred. No.:	84.8	Length:	152617
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	53.33%	Indels:	0
DB:	2	Gaps:	0

SEQ4-65T079 (1-15) x AC113881 (1-152617)

Oy 3 PheasnsSerleupProAsnAnlys 10
Db 117824 TTTAATAGCTTGCACACATATAA 117847

RESULT 9
AC021844

LOCUS Homo sapiens chromosome 5 clone RP11-15D23, WORKING DRAFT SEQUENCE,
4 unordered pieces.

AC021844 154455 bp DNA linear HTG 07-JUL-2000
AC021844 AC021844.4 GI:7534251
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 154455)
Waterston, R.H.

REFERENCE The sequence of Homo sapiens clone
AUTHORS Unpublished
TITLE 2 (bases 1 to 154455)
JOURNAL Waterston, R.H.

REFERENCE Direct Submission
AUTHORS Submitted (20-JAN-2000) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA

COMMENT On Apr 11, 2000 this sequence version replaced gi:7523912.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0015D23
----- Summary Statistics -----

Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 152020 bases at least Q40
Consensus quality: 152947 bases at least Q30

Consensus quality: 153535 bases at least Q20
Insert size: 16400; agarose-fp
Insert size: 15415; sum-of-contigs
Quality coverage: 5.45 in Q20 bases; agarose-fp
Quality coverage: 5.80 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 10635: contig of 10635 bp in length
* 10636 10735: gap of unknown length
* 10736 44171: contig of 33436 bp in length
* 44172 44271: gap of unknown length
* 44272 98542: contig of 54271 bp in length
* 98543 98642: gap of unknown length
* 98643 154455: contig of 55813 bp in length.

```

FEATURES
source
1. 154455
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-15D23"

misc_feature

1. 10635
/note="assembly_name:Contig3
clone_end:T7
vector_side:right"

misc_feature

10736. 44171
/note="assembly_name:Contig4"
44272. 98542
/note="assembly_name:Contig5"

misc_feature

98643. 154455
/note="assembly_name:Contig6
clone_end:SP6
vector_side:left"

BASE COUNT 48764 a 27957 c 28022 g 49405 t 307 others

ORIGIN

Alignment Scores:

Pred. No.:	85.7	Length:	154455
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	53.33%	Indels:	0
DB:	2	Gaps:	0

SEQ4-65T079 (1-15) x AC021844 (1-154455)

Oy 3 PheasnsSerleupProAsnAnlys 10

Db 56862 TTTAATAGCTTGCCTATACACAG 56885

RESULT 10
AC092328

LOCUS Homo sapiens chromosome 5 clone RP11-15D23, WORKING DRAFT SEQUENCE,
2 unordered pieces.
AC092328 AC092328.2 GI:15290443
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155394)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155394)

AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (03-JUN-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Aug 25, 2001 this sequence version replaced g1:14589517.
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

 Project Information
 Center Project Name: 412220
 Center Clone name: RPCI-11_15D23

 Summary Statistics
 Consensus quality: 153679 bases at least Q40
 Consensus quality: 153819 bases at least Q30
 Consensus quality: 153879 bases at least Q20
 Estimated insert size: 169360; agarose-*fp* estimation
 Estimated insert size: 155294; sum-of-*contigs* estimation
 Quality coverage: 12.32 in Q20 bases; agarose-*fp* estimation
 Quality coverage: 13.43 in Q20 bases; sum-of-*contigs* estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 *contigs*. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the *contigs* are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 25817: *contig* of 25817 bp in length
 * 25818 155394: gap of unknown length
 * 25918 155394: *contig* of 129477 bp in length.
 Location/Qualifiers
 1. 155394
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-15D23"
 /clone_1lb="RPCI human BAC library 11"
 BASE COUNT 46482 a 28012 c 28437 g 52363 t 100 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 86.1 Length: 155394
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 2 Gaps: 0
 SEQ4-65TO79 (1-15) x AC092328 (1-155394)
 QY 3 PheasSerLeuProAsnAsnLys 10
 ||||||||||||||||||||
 Db 57905 TTTAACCTCTTCCATAACACAG 57928
 RESULT 11
 AL390718
 DEFINITION Human DNA sequence from clone Rp11-502H18 on chromosome 1, complete
 sequence.
 ACCESSION AL390718
 VERSION AL390718.11 GI:16972947
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 176123)
 AUTHORS Wallis J.
 TITLE Direct Submission
 JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,

COMMENT Cambridgehire, CB10 1SA, UK. E-mail enquiries: humanyes@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Nov 16, 2001 this sequence version replaced g1:16444706.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone *contigs* of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
 RP11-502H18 is from the library RPCI-11.2 constructed by the group of Pletter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-502H18. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true left end of clone RP11-502H18 is at 1 in this sequence. The true left end of clone RP5-1180C10 is at 15124 in this sequence.
 Location/Qualifiers
 1. 176123
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-502H18"
 /clone_1lb="RPCI-11.2"
 BASE COUNT 49138 a 37981 c 39880 g 49124 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 95.9 Length: 176123
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 9 Gaps: 0
 SEQ4-65TO79 (1-15) x AL390718 (1-176123)
 QY 5 SerLeuProAsnAsnLysAlaSer 12
 ||||||||||||||||||||
 Db 70666 TCTCTCCCAATACACACATCA 70689
 RESULT 12
 AC025445
 LOCUS AC025445/C 184635 bp DNA linear PRI 06-SEP-2001
 DEFINITION Homo sapiens chromosome 5 clone CTD-2045017, complete sequence.
 ACCESSION AC025445
 VERSION AC025445.5 GI:15451688
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 184635)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission

```

JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 184635)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (09-MAR-2000) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3 (bases 1 to 184635)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Submitted (04-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
REFERENCE    4 (bases 1 to 184635)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Submitted (06-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Sep 6, 2001 this sequence version replaced gi:10567849.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www.sngc.stanford.edu
              Quality: Phrap Quality >=40 99.4% of Sequence:
              Estimated Total Number of Errors is 0.8.
              SRS Content:
              SHGC-15011 GI7012.
FEATURES     Location/Qualifiers
             source      1..184635
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="5"
                       /clone="CTD-2049017"
BASE COUNT   62006 a 33523 c 33180 g 55926 t
ORIGIN
Alignment Scores:
Pred. No.:      99.9      Length:      184635
Score:          8.00      Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    53.33% Indels: 0
DB:             9 Gaps: 0
SEQ4-65TO79 (1-15) x AC025445 (1-184635)
QY           3 pheasnSerleuProasnAnlys 10
             |||||||||||||||||||
Db 77793 TTTACTCTTTCCTATACACAG 77770

RESULT 13
AC068609/c 192263 bp DNA linear ROD 27-OCT-2001
LOCUS      Mus musculus chromosome 5 clone RP23-98L5 strain C57BL6/J, complete
sequence.
ACCESSION  AC068609
VERSION    AC068609.3 GI:16506406
KEYWORDS   HTG.
SOURCE     Mus musculus.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 192263)
AUTHORS    Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Ho,S.-L., Idol,J.R., Karins,E., Latic,P., Lee-Jin,S.-Q.,
            LegaSpi,R., Maduro,O.L., Maduro,V.B., Masteljo,C., Mastrian,S.D.,
            McLoskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
            Stantipop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
            Tsurgou,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Zhang,L.-H. and Green,E.D.
            NISC Comparative Sequencing Initiative
TITLE      JOURNAL
            Unpublished
REFERENCE  2 (bases 1 to 192263)

```

```

AUTHORS      Green,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (05-MAY-2000) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE    3 (bases 1 to 192263)
AUTHORS      Green,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT      On Oct 27, 2001 this sequence version replaced gi:10198311.
              ----- Genome Center
              Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc.mouse@nih.gov
              ----- Project Information
              Center project name: xv
              Center clone name: 098L05

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
all regions were covered by at least one plasmid subclone
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the Features section.
FEATURES     Location/Qualifiers
             source      1..192263
                       /organism="Mus musculus"
                       /strain="C57BL6/J"
                       /db_xref="taxon:10090"
                       /chromosome="5"
                       /clone="RP23-98L5"
                       /clone_lib="RPCT mouse BAC library 23"
             misc_feature 1..6259
                       /note="Clone overlaps with GenBank Accession Number
                       AC068665 (nucleotides 199679-205937) clone RP23-426K16
                       (center project name xv)"
             misc_feature 160265..160266
                       /note="bacterial transposon excised; IS1 element sequence
                       can be found in GenBank Accession Number AF000112.1
                       nucleotides 9258-10025; 768 bp"
BASE COUNT   55527 a 39492 c 40025 g 57219 t
ORIGIN
Alignment Scores:
Pred. No.:      103      Length:      192263
Score:          8.00      Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    53.33% Indels: 0
DB:             10 Gaps: 0
SEQ4-65TO79 (1-15) x AC068609 (1-192263)
QY           5 SerleuProasnAnlysaser 12
             |||||||||||||||||||
Db 164697 TCACCTACCAACAATAGCATCA 164674

Search completed: January 28, 2003, 12:16:35
Job time : 1591.67 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:32:26 ; Search time 344.333 Seconds
(without alignments)
98.103 Million cell updates/sec

Title: SEQ4-65TO79
Perfect score: 15
Sequence: 1 REFNSLPNNKASDTR 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/BORIN682/funat_23012003_130140_7866/app.query.fasta.1.597
-DB=N.Geneseq_101002 -OFMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCCALIGN=200 -THR.SCORE=quality -THR.MIN=8 -ALIGN=50 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=BORIN682.QCGN.1.1_187_@unat_23012003_130140_7866 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N.Geneseq_101002:.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
No matches found					

Search completed: January 28, 2003, 10:56:27
Job time : 344.333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:36:45 ; Search time 1270.67 seconds

(without alignments)
191.185 Million cell updates/sec

Title: SEQ4-65T079

Perfect score: 15

Sequence: 1 REFNSLPNKKASDPT 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 8

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: listing first 1000 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cg21.1/SPRO_sepool/BORIN682/rnatc_23012003_130141_7880/app_query.fasta.1.597
-DB=EST -QWMT=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LDOEXT=0
-ONITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000
-DOCALLIGN=200 -THR.SCORE=quality -THR.MIN=8 -ALIGN=50 -MODE=LOCAL -OUTFMT=pro
-NORMEXT -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=BORIN682 @CGN 1.1.1349 @runat.23012003_130141_7880 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estlro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	53.3	270	14	BQ987313
2	8	53.3	282	14	BQ986290
3	8	53.3	343	14	BQ983106
4	8	53.3	501	14	BQ010453
5	8	53.3	597	17	FR0048612
6	8	53.3	763	17	AZ698405

ALIGNMENTS

```

RESULT 1
BQ987313/C
LOCUS      BQ987313      270 bp      mRNA      linear      EST 21-AUG-2002
DEFINITION OQF1IN24.y9.ab1 QG_EFGHU lettuce serritola Lactuca sativa cDNA clone
ACCESSION  BQ987313
VERSION    BQ987313.1 GI:22404838
KEYWORDS   EST
SOURCE     Lactuca sativa.
ORGANISM   Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
            Lactuca.
            1 (bases 1 to 270)
            Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison,
            P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Churchill,S., Jackson,L. and Bradford,K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
            Unpublished (2002)
            Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig126, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QCF11 row: N column: 24.

FEATURES             source
     source
     1..270
     /organism="Lactuca sativa"
     /cultivar="L.serritola"
     /db_xref="taxon:4236"
     /clone="QCF1IN24"
     /clone_11b="QG_EFGHU lettuce serritola"
     /lab_host="E.coli"
     /note="Vector: pBRCDNASFLAB: The library was constructed
     from 10 different sources of RNA from a single genotype.
     Separate cDNAs were generated using primers that
     incorporated unique 5' and 3' tags to distinguish each
     source of RNA. cDNAs were then pooled, size-fractionated,
     directionally cloned into a custom medium-copy vector and
     transformations made with four size classes to minimize
     size bias. Details of each source of RNA and library
     construction can be obtained at http://cgpdb.ucdavis.edu/
     TAG_L1B-QG_EFGHU lettuce serritola
     TAG_T1TSUB-flowers environmental stress
     TAG_SEO-CGATGCGG"

BASE COUNT      73 a      40 c      73 g      84 t
ORIGIN

```

Alignment Scores:

Pred. No.: 51.8 Length: 270
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 14 Gaps: 0

SEQ4-65to79 (1-15) x BQ987313 (1-270)

OY 6 LeupProasnasnlYsAlaSerSer 13
 |||||||||||||||||||

Db 138 CTTCTAATATAATAAAGCTTCATCT 115

RESULT 2

BQ986290/c

LOCUS BQ986290 282 bp mRNA linear EST 21-AUG-2002
 DEFINITION OGE9B01.yg.ab1 OG_EFGHJ lettuce serriola lactuca sativa cDNA clone
 OGE9B01, mRNA sequence.

ACCESSION BQ986290

VERSION BQ986290.1 GI:22403815
 KEYWORDS EST.

SOURCE Lactuca sativa.

ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.

REFERENCE 1 (bases 1 to 282)

AUTHORS

Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
 ,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 unpublished (2002)

JOURNAL

COMMENT Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA.Contig126, see http://cgpbdb.ucdavis.edu/
 for details.

Plate: OGE9 row: B column: 01.

FEATURES

source

1..282
 Location/Qualifiers
 /organism="Lactuca sativa"
 /cultivar="L.serriola"
 /db_xref="taxon:4236"
 /clone="OGE9B01"
 /clone_1lb="OG_EFGHJ lettuce serriola"
 /lab_host="E.coli"
 /note="Vector: pBRCDNASFIAB: The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpbdb.ucdavis.edu/
 TAG_LIB-OG_EFGHJ lettuce serriola
 TAG_TISSUE=germinating seeds
 TAG_SEQ=TCGTCCGGG"

BASE COUNT 76 a 40 c 73 g 93 t
 ORIGIN

Alignment Scores:

Pred. No.: 54.2 Length: 282
 Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 14 Gaps: 0

SEQ4-65to79 (1-15) x BQ986290 (1-282)

OY 6 LeupProasnasnlYsAlaSerSer 13
 |||||||||||||||||||

Db 137 CTTCTAATATAATAAAGCTTCATCT 114

RESULT 3

BQ983106/c

LOCUS BQ983106 343 bp mRNA linear EST 21-AUG-2002
 DEFINITION OGE18D17.yg.ab1 OG_EFGHJ lettuce serriola lactuca sativa cDNA clone
 OGE18D17, mRNA sequence.

ACCESSION BQ983106

VERSION BQ983106.1 GI:2240631
 KEYWORDS EST.

SOURCE Lactuca sativa.

ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.

REFERENCE 1 (bases 1 to 343)

AUTHORS

Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
 ,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 unpublished (2002)

JOURNAL

COMMENT Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA.Contig126, see http://cgpbdb.ucdavis.edu/
 for details.

Plate: OGE18 row: D column: 17.

FEATURES

source

1..343
 Location/Qualifiers
 /organism="Lactuca sativa"
 /cultivar="L.serriola"
 /db_xref="taxon:4236"
 /clone="OGE18D17"
 /clone_1lb="OG_EFGHJ lettuce serriola"
 /lab_host="E.coli"
 /note="Vector: pBRCDNASFIAB: The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpbdb.ucdavis.edu/
 TAG_LIB-OG_EFGHJ lettuce serriola
 TAG_TISSUE=flowers environmental stress
 TAG_SEQ=CGAATGCGGCG"

BASE COUNT 96 a 48 c 82 g 117 t
 ORIGIN

Alignment Scores:

Pred. No.: 66.6 Length: 343
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.33% Indels: 0
 DB: 14 Gaps: 0

6 LeuProAsnAsnLysAlaSerSer 13
|||||

Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhaoc@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 240 Row: L Column: 10

Seq primer: SP6

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..763

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-240L10"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 246 a 160 c 148 g 209 t

ORIGIN

Alignment Scores:

Pred. No.:	154	Length:	763
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	53.33%	Indels:	0
DB:	17	Gaps:	0

SEQ4-65to79 (1-15) x AZ698405 (1-763)

QY 5 SerLeuProAsnAsnLysAlaSer 12

DB 281 TCACCTACCAACATATAGGCATCA 304

Search completed: January 28, 2003, 13:16:53
Job time : 1271.67 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:38:10 ; Search time 54 Seconds
(without alignments)
85.188 Million cell updates/sec

Title: SEQ4-65TO79
Perfect score: 15
Sequence: 1 REFNSLPNNKASDPT 15

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=x1h
-O=/cgn2_1/USPTO_SPOOL/BORIN682/rumat_23012003_130141_7891/app-query.fasta.1.597
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=olip2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALLIGN=200 -THR.SCORE=quality -THR.MIN=8 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=BORIN682_GCGN1.1_25_@rumat_23012003_130141_7891 -NCPU=6 -ICPU=3
-NO_XLUPX -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/lna/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/lna/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB	ID	Description

No matches found						

Search completed: January 28, 2003, 13:19:48
Job time : 54 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

Search completed: January 28, 2003, 13:23:11
Job time : 63 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 28, 2003, 10:39:05 ; Search time 63 Seconds
(without alignments)

106.968 Million cell updates/sec

Title: SEQ4-65TO79

Perfect score: 15
Sequence: 1 REFNSLPNNKASDTR 15

Scoring table: OLIGO

Xgapop	60.0	Xgapext	60.0
Ygapop	60.0	Ygapext	60.0
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 396772 seqs, 224632407 residues

Word size: 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/BORIN682/umat_23012003_130142_7908/app.query.fasta_1.597
-DB=Published_Applications_NA -QMT=fastlap -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=8
-ALIGN=50 -MODE=LOCAL -OUTFM=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USRR=BORIN682 @cgn_1_1_33 @umat_23012003_130142_7908
-KCPU=6 -ICPU=3 -NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
No.						

No matches found

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 ; Search time 127 Seconds
(without alignments)
15.738 Million cell updates/sec

Title: SEQ1-65TO79

Perfect score: 15
Sequence: 1 SLRSAHLAQSIILSG 15

Scoring table: OLIGO
Gapex 60.0, Capext 60.0

Searched: 908470 seqs, 133250620 residues

Word size: 8

Total number of hits satisfying chosen parameters: 29

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database:

A.Geneseq-101002:*

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	237	23 ABB07780	E coli mutant heat
2	15	100.0	240	19 AAW65074	E. coli Lr-A prote
3	15	100.0	240	23 ABB07778	E coli heat-labile
4	15	100.0	258	6 AAP50190	Sequence encoded b
5	15	100.0	258	6 AAP50191	Sequence encoded b
6	15	100.0	259	21 AAY96646	Plant-optimized E.
7	15	100.0	259	21 AAY96647	Synthetic E. coli
8	15	100.0	259	21 AAY96648	Plant-optimized E.
9	15	100.0	259	21 AAY96650	Plant-optimized E.
10	15	100.0	370	23 ABB07785	E coli mutant heat

11	15	100.0	373	23 ABB07784	E coli heat-labile
12	15	100.0	380	22 AAU00507	E. coli heat-labil
13	15	100.0	382	22 AAU00506	E. coli heat-labil
14	10	66.7	236	14 AAR44016	"Lys-63" E.coli he
15	10	66.7	236	14 AAR44017	"Lys-97" E.coli he
16	10	66.7	236	14 AAR44018	"Tyr-97" E.coli he
17	10	66.7	236	14 AAR44019	"Glu-107" E.coli h
18	10	66.7	236	14 AAR44020	"Lys-104" E.coli h
19	10	66.7	236	14 AAR44021	"Asp-104" E.coli h
20	10	66.7	236	14 AAR44022	"Ser-104" E.coli h
21	10	66.7	236	14 AAR44023	"Ser-106" E.coli h
22	10	66.7	236	14 AAR44024	"Glu-114" E.coli h
23	10	66.7	236	14 AAR44025	"Lys-114" E.coli h
24	10	66.7	236	14 AAR38728	E.coli heat labile
25	10	66.7	236	14 AAR38730	"Asp-53" E.coli he
26	10	66.7	236	14 AAR38731	"Glu-53" E.coli he
27	10	66.7	236	14 AAR38732	"Tyr-53" E.coli he
28	10	66.7	237	20 AAW67772	E. coli heat labil
29	10	66.7	254	22 AAU14105	Peptide sequence f

ALIGNMENTS

RESULT 1
ABB07780
ID ABB07780 standard: protein; 237 AA.
AC ABB07780:
XX
XX
XX
DT 17-JUN-2002 (first entry)
XX
XX
DE E coli mutant heat-labile toxin (mLT) A subunit fragment.
KW LT: heat-labile toxin; cholera toxin; CT; recombinant; adjuvant;
KW vaccine; mutant; mLT.
XX
XX
OS Escherichia coli.
XX
XX
PN JP2002051779-A.
XX
PD 19-FEB-2002.
XX
PF 07-AUG-2000; 2000JP-0238740.
XX
PR 07-AUG-2000; 2000JP-0238740.
XX
XX (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.
PA (HGET) HIGETA SHOUY KK.
PA (FUJI-) FUJITA GAKUEN.
XX
XX WPI; 2002-299402/34.
XX
XX
PT Preparation of a protein having 1A5B structure -
PS Claim 6; Fig 3; 27pp; Japanese.
XX
XX The invention relates to a gene encoding a protein having a subunit
CC structure of 1A5B in which the DNA sequence encoding each signal is
CC deleted from the A subunit gene and the B subunit gene and they are
CC combined tandemly in the order of (B subunit gene)-(SD sequence gene)
CC (A subunit gene). A method is provided for the preparation of a protein
CC having a subunit structure of 1A5B in which the above DNA is connected to
CC a vector expressible in Brevibacillus chosinensis and Brevibacillus
CC chosinensis is transformed by said vector and said transformant is
CC cultured. The protein can be used in the preparation of an adjuvant for
CC vaccine. The present sequence represents the E. coli mutant heat-labile
CC toxin (mLT) A subunit fragment, used for constructing a protein with the
CC subunit structure of 1A5B.
XX
XX
SQ Sequence 237 AA:
Query Match 100.0%; Score 15; DB 23; Length 237;

Best Local Similarity 100.0%; Pred. No. 6.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAGOSTLSG 15
|||||
Db 65 SLRSAHLAGOSTLSG 79

RESULT 2

AAW65074
ID AAW65074 standard; protein; 240 AA.

AC AAW65074;

DT 11-SEP-1998 (first entry)

DE E. coli LT-A protein fragment.

XX Heat labile toxin subunit A; LT-A; mutant; immunogen; detoxification;
KW carrier; adjuvant; prevention; treatment; disease; diarrhoea; vaccine;
KM infection; enterotoxic.

OS Escherichia coli.

XX Key Location/Qualifiers
FH 1..241
FT Protein /note= "partial sequence"

PN WO9818928-A1.

PD 07-MAY-1998.

PF 30-OCT-1997; 97WO-IB01440.

PR 31-OCT-1996; 96GB-0022660.

PA (CHIR-) CHIRON SPA.

PI Giuliani MM, Pizsa M, Rappuoli R;

DR WPI: 1998-272223/24.

XX Muted Escherichia coli heat labile toxin subunit A - is
PT immunogenic and detoxified relative to wild-type, useful e.g. in
XX vaccines against E. coli enterotoxigenic strains and as an adjuvant
PS
XX Disclosure: Page -: 67pp; English.

CC This sequence represents a fragment of an Escherichia coli heat labile
CC toxin subunit A (LT-A). This protein is used in a method resulting in a
CC mutant LT-A protein which has the wild type Ala residue at position 72
CC changed to an Arg residue resulting in a toxin which retains its
CC immunogenicity but is detoxified. Detoxification is defined in the
CC specification as a reduction in toxicity relative to wild-type toxin,
CC such that any residual toxicity is sufficiently low to allow use as an
CC effective immunogenic composition in humans without significant side
CC effects. The protein can be combined with an acceptable carrier in
CC immunogenic compositions, optionally comprising an adjuvant and/or a
CC second immunogenic antigen. Such compositions can be administered to
CC prevent/treat disease in a subject e.g. traveller's diarrhoea in humans.
CC The protein or compositions are especially administered as vaccines
CC useful to prevent or treat infections by enterotoxigenic strain of
CC E. coli in mammals (especially humans)..

XX Sequence 240 AA;

Query Match 100.0%; Score 15; DB 19; Length 240;

Best Local Similarity 100.0%; Pred. No. 6.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAGOSTLSG 15
|||||
Db 65 SLRSAHLAGOSTLSG 79

RESULT 3

ABB07778
ID ABB07778 standard; protein; 240 AA.

XX ABB07778;

DT 17-JUN-2002 (first entry)

DE E coli heat-labile toxin (LT) A subunit fragment.

XX LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant;
KW vaccine.

OS Escherichia coli.

PN JP2002051779-A.

PD 19-FEB-2002..

PF 07-AUG-2000; 2000JP-0238740.

PR 07-AUG-2000; 2000JP-0238740..

PA (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.
PA (HGET) HIGETA SHOYU KK.

PA (FUJI-) FUJITA GARDEN.

DR WPI: 2002-299402/34.

PT Preparation of a protein having 1A5B structure

PS Claim 5; Fig 1; 27pp; Japanese.

CC The invention relates to a gene encoding a protein having a subunit
CC structure of 1A5B in which the DNA sequence encoding each signal is
CC deleted from the A subunit gene and the B subunit gene and they are
CC combined tandemly in the order of (B subunit gene)-(SD sequence gene)-
CC (A subunit gene). A method is provided for the preparation of a protein
CC having a subunit structure of 1A5B in which the above DNA is connected to
CC a vector expressible in Brevibacillus chosinensis and Brevibacillus
CC chosinensis is transformed by said vector and said transformant is
CC cultured. The protein can be used in the preparation of an adjuvant for
CC vaccine. The present sequence represents the E. coli heat-labile toxin
CC (LT) A subunit fragment, used for constructing a protein with the subunit
CC structure of 1A5B.

XX Sequence 240 AA;

Query Match 100.0%; Score 15; DB 23; Length 240;

Best Local Similarity 100.0%; Pred. No. 6.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAGOSTLSG 15
|||||
Db 65 SLRSAHLAGOSTLSG 79

RESULT 4

AAP50190
ID AAP50190 standard; protein; 258 AA.

AC AAP50190;

DT 30-OCT-1991 (first entry)

DE Sequence encoded by the pig scours heat labile toxin (LT) LTA gene.

XX pig scours vaccine; toxin; diarrhoea.

OS E.coli NCIB 11932.

XX

```

PN  EP145486-A.
XX
PD  19-JUN-1985.
XX
PF  12-DEC-1984; 84EP-0308620.
XX
PR  12-DEC-1983; 83GB-0033131.
XX
PA  (GLAX ) GLAXO GROUP LTD.
XX
PI  Hayes MV, Harford S, Ross GW;
XX
DR  MPI; 1985-148358/25.
XX
DR  N-PSDB; AAN50205.
XX
PT  New toxoid as inactivated form of toxin for use in vaccines - is
XX  obtd. from organism transformed by gene
XX
PS  Disclosure: Fig 1; 61pp; English.
XX
CC  AAN50205 is the gene sequence of the natural LTA gene. The LTA gene of
XX  the site directed mutant SDM1 (see AAN50206) is inactive. The
XX  inventors claim a vaccine prepn. active against pig scours which
XX  contains an inactivated LTA component, together with additional K88
XX  antigens opt. with whole cells comprising the antigens or contg. the
XX  inactivated LTA.
XX
SQ  Sequence 258 AA;

Query Match          100.0%; Score 15; DB 6; Length 258;
Best Local Similarity 100.0%; Pred. No. 6.6e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 SLRSAHLAQSIILSG 15
    |||||
DB  83 SLRSAHLAQSIILSG 97

RESULT 5
AAP50191
ID  AAP50191 standard; Protein; 258 AA.
XX
AC  AAP50191;
XX
DT  30-OCT-1991 (first entry)
XX
DE  Sequence encoded by the pig scours heat labile toxin (LT) LTA gene
XX  of the site directed mutant SDM1.
XX
KM  Pig scours vaccine; toxin; diarrhoea.
XX
OS  E.coli NCIB 11932.
XX
FH  Key Location/Qualifiers
FT  Misc-difference 79 /note= "Ser in native SQ"
XX
XX  EP145486-A.
XX
XX  19-JUN-1985.
XX
XX
PF  12-DEC-1984; 84EP-0308620.
XX
PR  12-DEC-1983; 83GB-0033131.
XX
PA  (GLAX ) GLAXO GROUP LTD.
XX
PI  Hayes MV, Harford S, Ross GW;
XX
DR  MPI; 1985-148358/25.
XX
DR  N-PSDB; AAN50206.
XX
PT  New toxoid as inactivated form of toxin for use in vaccines - is

```

```

PT  obtd. from organism transformed by gene
XX
XX  Example; Fig 2; 61pp; English.
XX
CC  AAN50205 is the gene sequence of the natural LTA gene. The LTA gene of
XX  the site directed mutant SDM1 (see AAN50206) is inactive. The
XX  inventors claim a vaccine prepn. active against pig scours which
XX  contains an inactivated LTA component, together with additional K88
XX  antigens opt. with whole cells comprising the antigens or contg. the
XX  inactivated LTA.
XX
SQ  Sequence 258 AA;

Query Match          100.0%; Score 15; DB 6; Length 258;
Best Local Similarity 100.0%; Pred. No. 6.6e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 SLRSAHLAQSIILSG 15
    |||||
DB  83 SLRSAHLAQSIILSG 97

RESULT 6
AA996646
ID  AA996646 standard; Protein; 259 AA.
XX
AC  AA996646;
XX
DT  26-SEP-2000 (first entry)
XX
DE  Plant-optimized E. coli heat labile toxin A subunit.
XX
KM  Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
XX  adjuvant; anti-bacterial.
XX
OS  Escherichia coli.
XX
XX  Synthetic.
XX
FH  Key Location/Qualifiers
FT  Peptide 1..18
FT  Protein /label= signal_peptide
XX  19..259
XX  /label= mature_protein
XX
PN  MO200037609-A2.
XX
XX  29-JUN-2000.
XX
PF  22-DEC-1999; 99WO-US30747.
XX
PR  22-DEC-1998; 98US-0113507.
XX
PA  (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX  (MASO/) MASON H S.
XX  (ARNT/) ARNTZEN C J.
XX
PI  Mason HS, Arntzen CJ;
XX
XX  MPI; 2000-442653/38.
XX
XX  N-PSDB; AAA51106.
XX
PT  New polynucleotides encoding LT-A or CT-A polypeptides for the
XX  transformation of plant cells, useful in immunogenic compositions to
XX  elicit immune responses in animals
XX
PS  Example 1; Fig 1; 103pp; English.
XX
XX  This synthetic Escherichia coli heat-labile toxin (LT) A subunit (LT-A)
XX  is encoded by a plant-codon optimized cDNA. The cDNA sequence contains
XX  plant-preferred codons and eliminates sequence motifs associated with
XX  spurious mRNA processing. A single codon insertion (GTC encoding valine)
XX  was made to accommodate the creation of a NcoI restriction site around the
XX  initiator methionine codon. Novel polynucleotides encode a mutant LT-A

```


PT transformation of plant cells, useful in immunogenic compositions to
 PT elicit immune responses in animals
 XX
 PS Example 3; Page -: 103pp: English.
 CC This is mutant R192G Escherichia coli heat-labile toxin (LT) A subunit
 CC (LT-A). The wild-type arginine was replaced with glycine at residue 192
 CC of the mature protein, which was caused by a codon change of TCC to AAG
 CC in the coding sequence. The sequence contains plant-preferred codons and
 CC eliminates sequence motifs associated with spurious mRNA processing. A
 CC single codon insertion (GTG encoding valine) was made to accommodate the
 CC creation of a NcoI restriction site around the initiator methionine
 CC codon. Novel polynucleotides encode a mutant LT-A polypeptide or a
 CC mutant Vibrio cholerae cholera toxin (CT) A subunit (CT-A) polypeptide,
 CC which have reduced enzyme activity as compared to the wild-type LT-A or
 CC CT-A polypeptide and where at least one of the codons is altered to a
 CC plant preferred codon. The polynucleotide further comprises a nucleic
 CC acid sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The
 CC polynucleotides are useful for the transformation of plant cells for the
 CC production of transgenic plants to produce edible vaccines, especially
 CC oral vaccines in transgenic plants for the prophylactic or therapeutic
 CC treatment against E. coli or V. cholerae. The mutant polypeptides are
 CC also useful as adjuvants.
 CC NB: This sequence does not appear in the specification, it was made from
 CC the wild type sequence shown in AAY96647, which appears in Figure 1.
 SO Sequence 259 AA:
 Query Match 100.0%; Score 15; DB 21; Length 259;
 Best Local Similarity 100.0%; Pred. No. 6.7e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SRSAAHLAAGSILSG 15
 Db 84 SRSAAHLAAGSILSG 98
 RESULT 9
 AAY96650
 ID AAY96650 standard; Protein: 259 AA.
 AC AAY96650:
 XX
 DT 26-SEP-2000 (first entry)
 XX
 DE Plant-optimized E. coli LT-A A63K+R192G mutant.
 XX
 KW Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
 KW adjuvant; anti-bacterial; A63K+R192G; double mutant.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FH Peptide 1..19
 FT /label= signal_peptide
 FT Protein 20..259
 FT /label= mature_protein
 FT Misc-difference 62
 FT /label= S63K
 FT /note= "the wild type serine is replaced by lysine
 FT at position 63 of the mature sequence"
 FT Misc-difference 211
 FT /label= R192G
 FT /note= "the wild type arginine is replaced by glycine
 FT at position 192 of the mature sequence"
 FT
 FT
 PN WO200037609-A2.
 XX
 XX 29-JUN-2000.
 PD
 XX 22-DEC-1999; 99WO-US30747.
 XX

PR 22-DEC-1998; 98US-0113507.
 XX
 PA (BOYC-) BOYCE THOMPSON INST PLANT RES.
 PA (MASO/) MASON H S.
 PA (ARNT/) ARNTZEN C J.
 PI Mason HS, Arntzen CJ;
 DR WPI; 2000-442653/38.
 XX
 PT New polynucleotides encoding LT-A or CT-A polypeptides for the
 PT transformation of plant cells, useful in immunogenic compositions to
 PT elicit immune responses in animals
 XX
 PS Example 5; Page -: 103pp: English.
 CC This is Escherichia coli heat-labile toxin (LT) A subunit (LT-A) double
 CC mutant S63K+R192G. The wild-type serine was replaced with lysine at
 CC residue 63 and wild-type arginine was replaced with glycine at residue
 CC 192 of the mature protein. The coding sequence contains plant-preferred
 CC codons and eliminates sequence motifs associated with spurious mRNA
 CC processing. A single codon insertion (GTG encoding valine) was made to
 CC accommodate the creation of a NcoI restriction site around the initiator
 CC methionine codon. Novel polynucleotides encode a mutant LT-A polypeptide
 CC or a mutant Vibrio cholerae cholera toxin (CT) A subunit (CT-A)
 CC polypeptide, which have reduced enzyme activity as compared to the
 CC wild-type LT-A or CT-A polypeptide and where at least one of the codons
 CC is altered to a plant preferred codon. The polynucleotide further
 CC comprises a nucleic acid sequence encoding LT B subunit (LT-B) or a CT B
 CC subunit (CT-B). The polynucleotides are useful for the transformation of
 CC plant cells for the production of transgenic plants to produce edible
 CC vaccines, especially oral vaccines in transgenic plants for the
 CC prophylactic or therapeutic treatment against E. coli or V. cholerae.
 CC The mutant polypeptides are also useful as adjuvants.
 CC Note: This sequence does not appear in the specification. It was
 CC constructed from the wild type LT-A shown in AAY96646 which is given
 CC in Figure 1 of the specification.
 SO Sequence 259 AA:
 Query Match 100.0%; Score 15; DB 21; Length 259;
 Best Local Similarity 100.0%; Pred. No. 6.7e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SRSAAHLAAGSILSG 15
 Db 84 SRSAAHLAAGSILSG 98
 RESULT 10
 ABB07785
 ID ABB07785 standard; Protein: 370 AA.
 AC ABB07785:
 XX
 DT 17-JUN-2002 (first entry)
 XX
 DE E coli mutant heat-labile toxin (mLT) 5B-SD-1A protein.
 XX
 KW LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant;
 KW vaccine; mutant; mLT.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FH Peptide 1..103
 FT /note= "LT 5B subunit"
 FT Peptide 104..370
 FT /note= "LT 1A subunit"
 XX
 XX JP2002051779-A.
 PN
 XX 19-FEB-2002.
 XX

```

XX 07-AUG-2000; 2000JP-0238740.
PF 07-AUG-2000; 2000JP-0238740.
PR 07-AUG-2000; 2000JP-0238740.
XX
XX (DOKU-) DOKURITSU GYOSAI HOJIN NOGYO SEIBUTSU SH.
PA (HGET) HIGETA SHOYU KK.
PA (FUJI-) FUJITA GAKUEN.
XX
XX WPI: 2002-299402/34.
DR N-PSDB; ABL40640.
XX
XX Preparation of a protein having 1A5B structure -
PT Disclosure; Fig 9-10; 27pp; Japanese.
XX
XX The invention relates to a gene encoding a protein having a subunit
CC structure of 1A5B in which the DNA sequence encoding each signal is
CC deleted from the A subunit gene and the B subunit gene and they are
CC combined tandemly in the order of (B subunit gene)-(SD sequence gene)-
CC (A subunit gene). A method is provided for the preparation of a protein
CC having a subunit structure of 1A5B in which the above DNA is connected to
CC a vector expressible in Brevibacillus chosinensis and Brevibacillus
CC chosinensis is transformed by said vector and said transformant is
CC cultured. The protein can be used in the preparation of an adjuvant for
CC vaccine. The present sequence represents the E. coli mutant heat-labile
CC toxin (MLT) 5B-SD-1A protein.
XX
SQ Sequence 370 AA:
Query Match 100.0%; Score 15; DB 23; Length 370;
Best Local Similarity 100.0%; Pred. NO. 9.1e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SLRSAHLAGSILSG 15
Db 198 SLRSAHLAGSILSG 212

RESULT 11
ABBO7784
ID ABBO7784 standard; Protein; 373 AA.
XX
XX ABB07784;
AC 17-JUN-2002 (first entry)
DE E coli heat-labile toxin (LT) 5B-SD-1A protein.
XX
XX LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant;
KW vaccine.
XX
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FT Peptide 1..103
FT /note= "LT 5B subunit"
FT 104..373
FT /note= "LT 1A subunit"
FT
XX
XX JP2002051779-A.
PN
XX
XX 19-FEB-2002.
PD
XX
XX 07-AUG-2000; 2000JP-0238740.
PF
XX
XX 07-AUG-2000; 2000JP-0238740.
PR
XX
XX (DOKU-) DOKURITSU GYOSAI HOJIN NOGYO SEIBUTSU SH.
PA (HGET) HIGETA SHOYU KK.
PA (FUJI-) FUJITA GAKUEN.
XX
XX WPI: 2002-299402/34.
DR

```

```

DR N-PSDB; ABL40639.
XX
XX Preparation of a protein having 1A5B structure -
PT Disclosure; Fig 7-8; 27pp; Japanese.
XX
XX The invention relates to a gene encoding a protein having a subunit
CC structure of 1A5B in which the DNA sequence encoding each signal is
CC deleted from the A subunit gene and the B subunit gene and they are
CC combined tandemly in the order of (B subunit gene)-(SD sequence gene)-
CC (A subunit gene). A method is provided for the preparation of a protein
CC having a subunit structure of 1A5B in which the above DNA is connected to
CC a vector expressible in Brevibacillus chosinensis and Brevibacillus
CC chosinensis is transformed by said vector and said transformant is
CC cultured. The protein can be used in the preparation of an adjuvant for
CC vaccine. The present sequence represents the E. coli heat-labile toxin
CC (LT) 5B-SD-1A protein.
XX
SQ Sequence 373 AA:
Query Match 100.0%; Score 15; DB 23; Length 373;
Best Local Similarity 100.0%; Pred. NO. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SLRSAHLAGSILSG 15
Db 198 SLRSAHLAGSILSG 212

RESULT 12
AAU00507
ID AAU00507 standard; Protein; 380 AA.
XX
XX AAU00507;
AC 29-AUG-2001 (first entry)
DE E. coli heat-labile enterotoxin (LT) mutant LTdel110/112.
XX
XX Heat-labile enterotoxin; LT; LT63Y; LTdel110/112; mutant;
KW detoxified and immunologically active protein; ADP-ribosylation; Gs;
KW endotoxin; diarrhoea; mucin.
XX
XX Escherichia coli strain K88ac.
OS
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FT MISC_feature 7
FT /note= "Important residue for enzymatic activity"
FT
FT MISC_feature 44
FT /note= "Important residue for enzymatic activity"
FT
FT Region 58..72
FT /note= "Forms the NAD-binding site"
FT
FT MISC_feature 61
FT /note= "Important residue for enzymatic activity"
FT
FT MISC_difference 257
FT /note= "Encoded by TG"
FT
XX
XX MO200119998-A1.
PN
XX
XX 22-MAR-2001.
PD
XX
XX 15-SEP-1999; 99WO-KR00555.
PF
XX
XX 15-SEP-1999; 99WO-KR00555.
PR
XX
XX (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
PA
PA Park EJ, Kim JS, Chang J, Yum J, Chung S;
XX
XX WPI: 2001-281524/29.
DR
DR N-PSDB; AAS01506.
XX

```


PT New detoxified mutants of *Escherichia coli* heat-labile enterotoxin
 PT useful as vaccine for preventing and treating diarrhoea, and as adjuvant
 PT for antibody production
 PS Claim 6; Page 42-44; 48pp; English.
 XX
 CC The present sequence represents *Escherichia coli* heat-labile
 CC enterotoxin (LT) mutant LTdell110/112. LRS63Y (AAU00506) and LTdell110/112
 CC are two novel detoxified and immunologically active proteins (LT
 CC mutants) derived by site-directed mutagenesis of the A1 subunit of wild
 CC type LT. The substitution of Ser to Tyr at position 63 in LRS63Y blocks
 CC NAD-binding. Deletion of Glu residues at positions 110 and 112 in
 CC LTdell110/112 eliminate the enzymatic activity of LT. The A1 subunit of
 CC wild type LT catalyses ADP-ribosylation of Gs, a GTP-binding protein that
 CC regulates cAMP levels. The resulting increase in cAMP is the cause of
 CC diarrhoea in humans and animals e.g. pigs. The mucosal immunogenicities
 CC of mutant heat-labile endotoxins LRS63Y and LTdell110/112 were tested.
 CC Groups of mice were immunised with LRS623Y or LTdell110/112. The control
 CC groups received phosphate buffered saline (PBS) alone. The serum and
 CC faecal antibody titres to LT were determined. The results showed that
 CC mice immunised with LRS63Y or LTdell110/112 contained high and
 CC comparable level of anti-LT antibodies in sera and faecal extracts
 CC compared with those immunised with wild-type LT. The LT mutants are
 CC useful as a vaccine for preventing and treating diarrhoea and as an
 CC adjuvant for antibody production.
 XX
 SQ Sequence 380 AA:
 Query Match 100.0%; Score 15; DB 22; Length 380;
 Best Local Similarity 100.0%; Pred. No. 9.4e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SLRSAPHLAQSIISLG 15
 DB 83 SLRSAPHLAQSIISLG 97
 RESULT 13
 AAU00506
 ID AAU00506 standard; Protein; 382 AA.
 AC AAU00506;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE E. coli heat-labile enterotoxin (LT) mutant LRS63Y.
 XX
 KW Heat-labile enterotoxin; LT; LRS63Y; LTdell110/112; mutant;
 KW detoxified and immunologically active protein; ADP-ribosylation; Gs;
 XX endotoxin; diarrhoea; mutela.
 XX
 OS *Escherichia coli* strain K88ac.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT MISC-feature 7
 FT MISC-feature /note= "Important residue for enzymatic activity"
 FT MISC-feature 44
 FT MISC-feature /note= "Important residue for enzymatic activity"
 FT Region 58..72
 FT MISC-feature /note= "Forms the NAD-binding site"
 FT MISC-feature 61
 FT MISC-feature /note= "Important residue for enzymatic activity"
 FT MISC-difference 63
 FT MISC-feature /note= "Substitution of wild type Ser to Tyr"
 FT MISC-feature 110
 FT MISC-feature /note= "Important residue for enzymatic activity"
 FT MISC-feature 112
 FT MISC-difference /note= "Important residue for enzymatic activity"
 FT MISC-difference 259
 FT /note= "Encoded by TG"
 XX
 PN WO200119998-A1.

XX
 PD 22-MAR-2001.
 XX
 PF 15-SEP-1999; 99WO-KR00555.
 XX
 PR 15-SEP-1999; 99WO-KR00555.
 XX
 PA (MOGA-) MOGAM BIOECHOLOGY RES INST.
 XX
 PI Park EJ, Kim JS, Chang J, Yum J, Chung S;
 XX
 DR MPI: 2001-281524/29.
 DR N-PSDB: AAS01505.
 XX
 PT New detoxified mutants of *Escherichia coli* heat-labile enterotoxin
 PT useful as vaccine for preventing and treating diarrhoea, and as adjuvant
 PT for antibody production
 PS Claim 2; Page 39-41; 48pp; English.
 XX
 CC The present sequence represents *Escherichia coli* heat-labile
 CC enterotoxin (LT) mutant LRS63Y. LRS63Y and LTdell110/112 (AAU00507)
 CC are two novel detoxified and immunologically active proteins (LT
 CC mutants) derived by site-directed mutagenesis of the A1 subunit of wild
 CC type LT. The substitution of Ser to Tyr at position 63 in LRS63Y blocks
 CC NAD-binding. Deletion of Glu residues at positions 110 and 112 in
 CC LTdell110/112 eliminate the enzymatic activity of LT. The A1 subunit of
 CC wild type LT catalyses ADP-ribosylation of Gs, a GTP-binding protein that
 CC regulates cAMP levels. The resulting increase in cAMP is the cause of
 CC diarrhoea in humans and animals e.g. pigs. The mucosal immunogenicities
 CC of mutant heat-labile endotoxins LRS63Y and LTdell110/112 were tested.
 CC Groups of mice were immunised with LRS623Y or LTdell110/112. The control
 CC groups received phosphate buffered saline (PBS) alone. The serum and
 CC faecal antibody titres to LT were determined. The results showed that
 CC mice immunised with LRS63Y or LTdell110/112 contained high and
 CC comparable level of anti-LT antibodies in sera and faecal extracts
 CC compared with those immunised with wild-type LT. The LT mutants are
 CC useful as a vaccine for preventing and treating diarrhoea and as an
 CC adjuvant for antibody production.
 XX
 SQ Sequence 382 AA:
 Query Match 100.0%; Score 15; DB 22; Length 382;
 Best Local Similarity 100.0%; Pred. No. 9.4e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SLRSAPHLAQSIISLG 15
 DB 83 SLRSAPHLAQSIISLG 97
 RESULT 14
 AAR44016
 ID AAR44016 standard; Protein; 236 AA.
 AC AAR44016;
 XX
 DT 08-DEC-1993 (first entry)
 XX
 DE "Lys-63" *E. coli* heat labile toxin subunit A.
 XX
 KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
 KW protomer A; site-directed mutagenesis; reduced toxicity;
 KW ADP-ribosyltransferase activity.
 XX
 OS *Escherichia coli*.
 OS
 FH Key Location/Qualifiers
 FT MISC-difference 62
 FT MISC-difference /note= "corresponds to position 63 in cholera toxin
 FT A subunit; wild-type Ser is substituted by
 FT Lys to reduce toxicity"
 XX

PN W09313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92WO-EP03016.
XX
XX 31-DEC-1991; 91IT-0MI3513.
XX
XX (BIOC-) BIOGINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX N-PSDB; AAQ51317.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
XX PT - useful as vaccines against infection by Vibrio cholerae and
XX PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
XX CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
XX CC et al, J.Biol. Chem., 259, 5037-5044 - see AAQ42768) was subjected to
XX CC site-directed mutagenesis. Certain mutations were found to reduce
XX CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
XX CC immunogenic, detoxified LT-A proteins and their use in vaccines to
XX CC protect against enterotoxigenic E.coli. The amino acid sequence of
XX CC this preferred detoxified mutain is not printed in the
XX CC specification but has been assembled from the full-length wild-type
XX CC sequence and the description given in the text. (N.B. Amino
XX CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 236 AA:

Query Match 66.7%; Score 10; DB 14; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQ 10
DB 64 SLRSAHLAQ 73

RESULT 15

AAR44017
ID AAR44017 standard; Protein; 236 AA.

XX AAR44017;

XX 08-DEC-1993 (first entry)

XX "Lys-97" E.coli heat labile toxin subunit A.

XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;

XX protomer A; site-directed mutagenesis; reduced toxicity;

XX ADP-ribosyltransferase activity.

XX Escherichia coli.

XX Key Location/Qualifiers

XX Misc-difference 93 /note= "corresponds to position 97 in cholera toxin
XX A subunit; wild-type Val is substituted by
XX Lys to reduce toxicity"

XX W09313202-A.

XX 08-JUL-1993.

XX 30-DEC-1992; 92WO-EP03016.

XX 31-DEC-1991; 91IT-0MI3513.

XX (BIOC-) BIOGINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Pizza M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX N-PSDB; AAQ51318.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
XX PT - useful as vaccines against infection by Vibrio cholerae and
XX PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.

XX The wild-type sequence coding for the A subunit of the heat labile
XX CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
XX CC et al, J.Biol. Chem., 259, 5037-5044 - see AAQ42768) was subjected to
XX CC site-directed mutagenesis. Certain mutations were found to reduce
XX CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
XX CC immunogenic, detoxified LT-A proteins and their use in vaccines to
XX CC protect against enterotoxigenic E.coli. The amino acid sequence of
XX CC this preferred detoxified mutain is not printed in the
XX CC specification but has been assembled from the full-length wild-type
XX CC sequence and the description given in the text. (N.B. Amino
XX CC acid numbering is based on the cholera toxin A subunit sequence).

SQ Sequence 236 AA:

Query Match 66.7%; Score 10; DB 14; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQ 10
DB 64 SLRSAHLAQ 73

RESULT 16

AAR44018
ID AAR44018 standard; Protein; 236 AA.

XX AAR44018;

XX 08-DEC-1993 (first entry)

XX "Tyr-97" E.coli heat labile toxin subunit A.

XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;

XX protomer A; site-directed mutagenesis; reduced toxicity;

XX ADP-ribosyltransferase activity.

XX Escherichia coli.

XX Key Location/Qualifiers

XX Misc-difference 93 /note= "corresponds to position 97 in cholera toxin
XX A subunit; wild-type Val is substituted by
XX Tyr to reduce toxicity"

XX W09313202-A.

XX 08-JUL-1993.

XX 30-DEC-1992; 92WO-EP03016.

XX 31-DEC-1991; 91IT-0MI3513.

XX (BIOC-) BIOGINE SCLAVO SPA.

XX Domenighini M, Hol W, Pizza M, Rappuoli R;

XX WPI; 1993-227320/28.

XX N-PSDB; AAQ51319.

```

XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of
CC this preferred detoxified muten is not printed in the
CC specification but has been assembled from the full-length wild-type
CC sequence and the description given in the text. (N.B. Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
XX Sequence 236 AA:
SQ
XX
XX Query Match 66.7%; Score 10; DB 14; Length 236;
XX Best Local Similarity 100.0%; Pred. No. 0.0091;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 SLRSAPHLAQ 10
XX |||||||||
XX DB 64 SLRSAPHLAQ 73
XX
XX RESULT 17
XX AAR44019
XX ID AAR44019 standard; Protein; 236 AA.
XX
XX AAR44019;
XX
XX 08-DEC-1993 (first entry)
XX
XX "Glu-107" E.coli heat labile toxin subunit A.
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
XX protomer A; site-directed mutagenesis; reduced toxicity;
XX ADP-ribosyltransferase activity.
XX
XX Escherichia coli.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 103
XX FT /note="corresponds to position 107 in cholera toxin
XX FT A subunit; wild-type His is substituted by
XX FT Glu to reduce toxicity"
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92WO-EP03016.
XX
XX 31-DEC-1991; 91IT-0M13513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Piazza M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX N-PSDB; AA051320.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX

```

```

CC The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of
CC this preferred detoxified muten is not printed in the
CC specification but has been assembled from the full-length wild-type
CC sequence and the description given in the text. (N.B. Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
XX Sequence 236 AA:
SQ
XX
XX Query Match 66.7%; Score 10; DB 14; Length 236;
XX Best Local Similarity 100.0%; Pred. No. 0.0091;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 SLRSAPHLAQ 10
XX |||||||||
XX DB 64 SLRSAPHLAQ 73
XX
XX RESULT 18
XX AAR44020
XX ID AAR44020 standard; Protein; 236 AA.
XX
XX AAR44020;
XX
XX 08-DEC-1993 (first entry)
XX
XX "Lys-104" E.coli heat labile toxin subunit A.
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
XX protomer A; site-directed mutagenesis; reduced toxicity;
XX ADP-ribosyltransferase activity.
XX
XX Escherichia coli.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 100
XX FT /note="corresponds to position 104 in cholera toxin
XX FT A subunit; wild-type Tyr is substituted by
XX FT Lys to reduce toxicity"
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92WO-EP03016.
XX
XX 31-DEC-1991; 91IT-0M13513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Piazza M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX N-PSDB; AA051321.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of

```

CC this preferred detoxified mutin is not printed in the
 CC specification but has been assembled from the full-length wild-type
 CC sequence and the description given in the text. (N.B. Amino
 CC acid numbering is based on the cholera toxin A subunit sequence).

XX Sequence 236 AA;

Query Match 66.7%; Score 10; DB 14; Length 236;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQ 10
 |||||
 Db 64 SLRSAHLAQ 73

RESULT 19

AAR44021 AAR44021 standard; Protein; 236 AA.

XX AAR44021;

DT 08-DEC-1993 (first entry)

XX "Asp-104" E.coli heat labile toxin subunit A.

XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;

KM protomer A; site-directed mutagenesis; reduced toxicity;

XX ADP-ribosyltransferase activity.

OS Escherichia coli.

XX Key Location/Qualifiers

FT Misc-difference 100 /note="corresponds to position 104 in cholera toxin
 FT A subunit; wild-type Tyr is substituted by
 FT Asp to reduce toxicity"

PN W09313202-A.

PD 08-JUL-1993.

XX 30-DEC-1992; 92WO-EP03016.

XX 31-DEC-1991; 91IT-OMI3513.

XX (BIOC-) BIOCINE SCLAVO SPA.

XX Domenighini M, Hol W, Pizza M, Rappuoli R;

PI WPI; 1993-227320/28.

DR N-PSDB; AA051322.

XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
 PT - useful as vaccines against infection by Vibrio cholerae and
 PT enterotoxin producing Escherichia coli

XX Claim 3; Fig 2 and Page 46; 60pp; English.

CC The wild-type sequence coding for the A subunit of the heat labile
 CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
 CC et al, J.Biol. Chem., 259, 5037-5044 - see AAQ42768) was subjected to
 CC site-directed mutagenesis. Certain mutations were found to reduce
 CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
 CC immunogenic, detoxified LT-A proteins and their use in vaccines to
 CC protect against enterotoxigenic E.coli. The amino acid sequence of
 CC this preferred detoxified mutin is not printed in the
 CC specification but has been assembled from the full-length wild-type
 CC sequence and the description given in the text. (N.B. Amino
 CC acid numbering is based on the cholera toxin A subunit sequence).

XX Sequence 236 AA;

Query Match 66.7%; Score 10; DB 14; Length 236;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQ 10
 |||||
 Db 64 SLRSAHLAQ 73

RESULT 20

AAR44022 AAR44022 standard; Protein; 236 AA.

XX AAR44022;

DT 08-DEC-1993 (first entry)

XX "Ser-104" E.coli heat labile toxin subunit A.

XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;

KM protomer A; site-directed mutagenesis; reduced toxicity;

XX ADP-ribosyltransferase activity.

OS Escherichia coli.

XX Key Location/Qualifiers

FT Misc-difference 100 /note="corresponds to position 104 in cholera toxin
 FT A subunit; wild-type Tyr is substituted by
 FT Ser to reduce toxicity"

PN W09313202-A.

PD 08-JUL-1993.

XX 30-DEC-1992; 92WO-EP03016.

XX 31-DEC-1991; 91IT-OMI3513.

XX (BIOC-) BIOCINE SCLAVO SPA.

XX Domenighini M, Hol W, Pizza M, Rappuoli R;

PI WPI; 1993-227320/28.

DR N-PSDB; AA051323.

XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
 PT - useful as vaccines against infection by Vibrio cholerae and
 PT enterotoxin producing Escherichia coli

XX Claim 3; Fig 2 and Page 46; 60pp; English.

CC The wild-type sequence coding for the A subunit of the heat labile
 CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
 CC et al, J.Biol. Chem., 259, 5037-5044 - see AAQ42768) was subjected to
 CC site-directed mutagenesis. Certain mutations were found to reduce
 CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
 CC immunogenic, detoxified LT-A proteins and their use in vaccines to
 CC protect against enterotoxigenic E.coli. The amino acid sequence of
 CC this preferred detoxified mutin is not printed in the
 CC specification but has been assembled from the full-length wild-type
 CC sequence and the description given in the text. (N.B. Amino
 CC acid numbering is based on the cholera toxin A subunit sequence).

XX Sequence 236 AA;

Query Match 66.7%; Score 10; DB 14; Length 236;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQ 10
 |||||
 Db 64 SLRSAHLAQ 73

```

RESULT 21
AAR44023
ID AAR44023 standard; Protein: 236 AA.
XX
XX
AC AAR44023:
XX
DT 08-DEC-1993 (first entry)
XX
DE "Ser-106" E.coli heat labile toxin subunit A.
XX
KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity.
XX
OS Escherichia coli.
XX
Key Location/Qualifiers
FH Misc-difference 102
FT /note="corresponds to position 106 in cholera toxin
FT A subunit; wild-type Pro is substituted by
FT Ser to reduce toxicity"
FT
XX
XX
XX MO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992; 92MO-EP03016.
XX
XX 31-DEC-1991; 91IT-OMI3513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Piza M, Rappuoli R;
XX
XX WPI: 1993-227320/28.
XX
XX N-PSDB: AA051324.
XX
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
XX - useful as vaccines against infection by Vibrio cholerae and
XX enterotoxin producing Escherichia coli
XX
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
XX toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
XX et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
XX site-directed mutagenesis. Certain mutations were found to reduce
XX toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
XX immunogenic, detoxified LT-A proteins and their use in vaccines to
XX protect against enterotoxigenic E.coli. The amino acid sequence of
XX this preferred detoxified mutetin is not printed in the
XX specification but has been assembled from the full-length wild-type
XX sequence and the description given in the text. (N.B. Amino
XX acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 236 AA:
XX
Query Match 66.7%; Score 10; DB 14; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 64 SLRSNHLAQ 73
QY 1 SLRSNHLAQ 10
IIIIIIIIII
AAR44024
ID AAR44024 standard; Protein: 236 AA.
XX
XX AAR44024:
XX

```

XX	08-DEC-1993	(first entry)	
DT			
XX	"Glu-114" E.coli heat labile toxin subunit A.		
DE			
XX	enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;		
KW	protomer A; site-directed mutagenesis; reduced toxicity;		
KW	ADP-ribosyltransferase activity.		
XX			
OS	Escherichia coli.		
XX			
XX	Key	Location/Qualifiers	
FT	Misc-difference 110	/note= "corresponds to position 114 in cholera toxin	
FT		A subunit; wild-type Ser is substituted by	
FT		Glu to reduce toxicity"	
XX			
PN	W09J13202-A.		
XX			
XX	08-JUL-1993.		
XX			
PF	30-DEC-1992;	92WO-EP03016.	
XX			
PR	31-DEC-1991;	91IT-0M13513.	
XX			
PA	(BIOC-) BIOGINE SCLAVO SPA.		
PI	Domenighini M, Hol W, Pizsa M, Rappuoli R;		
XX			
DR	WPI: 1993-227320/28.		
DR	N-PSDB; AA051325.		
XX			
PT	Immunogenic detoxified mutant cholera toxin and heat labile toxin		
PT	- useful as vaccines against infection by Vibrio cholerae and		
PT	enterotoxin producing Escherichia coli		
XX			
PS	Claim 3; Fig 2 and Page 46; 60pp; English.		
XX			
CC	The wild-type sequence coding for the A subunit of the heat labile		
CC	toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto		
CC	et al., J Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to		
CC	site-directed mutagenesis. Certain mutations were found to reduce		
CC	toxicity (see AA38730-32 and AA44016-R44025). The invention relates to		
CC	immunogenic, detoxified LT-A proteins and their use in vaccines to		
CC	protect against enterotoxigenic E.coli. The amino acid sequence of		
CC	this preferred detoxified mutain is not printed in the		
CC	specification but has been assembled from the full-length wild-type		
CC	sequence and the description given in the text. (N.B. Amino		
CC	acid numbering is based on the cholera toxin A subunit sequence).		
XX			
SQ	Sequence 236 AA;		
	Query Match	66.7%; Score 10; DB 14; Length 236;	
	Best Local Similarity 100.0%; Pred. No. 0.0091;		
	Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 SLSAHLAQ 10		
DB	64 SLSAHLAQ 73		
RESULT 23			
AA44025			
ID	AA44025 standard; Protein; 236 AA.		
XX			
AC	AA44025;		
XX			
DT	08-DEC-1993 (first entry)		
XX			
DE	"Lys-114" E.coli heat labile toxin subunit A.		
XX			
KW	enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;		
KW	protomer A; site-directed mutagenesis; reduced toxicity;		
XX			

```

KW ADP-ribosyltransferase activity.
XX Escherichia coli.
OS
XX Key Location/Qualifiers
FH Misc-difference 110 /note= "corresponds to position 114 in cholera toxin
FT A subunit; wild-type Ser is substituted by
FT Lys to reduce toxicity"
FT
XX
XX
XX WO9313202-A.
XX
XX PD 08-JUL-1993.
XX PF 30-DEC-1992; 92MO-EP03016.
XX PR 31-DEC-1991; 91IT-0M13513.
XX PA (BIOC-) BIOGINE SCLAVO SPA.
XX PI Domenighini M, Hol W, Piazza M, Rappuoli R;
XX WPI: 1993-227320/28.
XX DR N-PSDB; AA051326.
XX
XX PT Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
PT
XX
XX PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX CC The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AAR38730-32 and AAR4016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of
CC this preferred detoxified muten is not printed in the
CC specification but has been assembled from the full-length wild-type
CC sequence and the description given in the text. (N.B. Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
XX SQ Sequence 236 AA;

```

```

Query Match 66.7%; Score 10; DB 14; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 SLRSHLAGQ 10
   |||||||||
DB 64 SLRSHLAGQ 73

```

```

RESULT 24
AAR38728
ID AAR38728 standard; Protein; 236 AA.
XX
XX AC AAR38728;
XX
XX DT 08-DEC-1993 (first entry)
XX
XX DE E.coli heat labile toxin subunit A.
XX
XX KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
XX protomer A; site-directed mutagenesis; reduced toxicity;
XX ADP-ribosyltransferase activity.
XX
XX OS Escherichia coli.
XX
XX PN WO9313202-A.
XX
XX PD 08-JUL-1993.

```

```

XX
XX PF 30-DEC-1992; 92MO-EP03016.
XX
XX PR 31-DEC-1991; 91IT-0M13513.
XX
XX PA (BIOC-) BIOGINE SCLAVO SPA.
XX
XX PI Domenighini M, Hol W, Piazza M, Rappuoli R;
XX WPI: 1993-227320/28.
XX DR N-PSDB; AA042768.
XX
XX PT Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
PT
XX
XX PS Disclosure; Fig 2; 60pp; English.
XX
XX CC This is the sequence of the A subunit of the heat labile toxin (LT-A)
XX of a strain of E.coli known to affect humans. The sequence was
XX published by Yamamoto et al, J Biol. Chem., 259, 5037-5044.
XX Mutations at selected positions within this sequence have been found
XX to reduce toxicity (see AAR38730-R38732 and AAR4016-R44025). The
XX invention relates to such immunogenic, detoxified proteins and their
XX use in vaccines to protect against enterotoxigenic E.coli.
XX
XX SQ Sequence 236 AA;

```

```

Query Match 66.7%; Score 10; DB 14; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 SLRSHLAGQ 10
   |||||||||
DB 64 SLRSHLAGQ 73

```

```

RESULT 25
AAR38730
ID AAR38730 standard; Protein; 236 AA.
XX
XX AC AAR38730;
XX
XX DT 08-DEC-1993 (first entry)
XX
XX DE "Asp-53" E.coli heat labile toxin subunit A.
XX
XX KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
XX protomer A; site-directed mutagenesis; reduced toxicity;
XX ADP-ribosyltransferase activity.
XX
XX OS Escherichia coli.
XX
XX FH Key Location/Qualifiers
FH Misc-difference 52 /note= "corresponds to position 53 in cholera toxin
FT A subunit; wild-type Val is substituted by
FT Asp to reduce toxicity"
FT
XX
XX PN WO9313202-A.
XX
XX PD 08-JUL-1993.
XX
XX PF 30-DEC-1992; 92MO-EP03016.
XX
XX PR 31-DEC-1991; 91IT-0M13513.
XX
XX PA (BIOC-) BIOGINE SCLAVO SPA.
XX
XX PI Domenighini M, Hol W, Piazza M, Rappuoli R;
XX WPI: 1993-227320/28.
XX DR N-PSDB; AA051314.

```

```

XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
PS Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of
CC this preferred detoxified mutain is not printed in the
CC specification but has been assembled from the full-length wild-type
CC sequence and the description given in the text. (N.B. Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 236 AA:
Query Match 66.7%; Score 10; DB 14; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SLRSAHLAQ 10
DB 64 SLRSAHLAQ 73
RESULT 26
AAR38731
ID AAR38731 standard; Protein; 236 AA.
AC AAR38731:
XX
XX 08-DEC-1993 (first entry)
DE "Glu-53" E.coli heat labile toxin subunit A.
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KM ADP-ribosyltransferase activity.
XX
XX Escherichia coli.
OS
FH Key Location/Qualifiers
FT Misc-difference 52 /note="corresponds to position 53 in cholera toxin
FT A subunit; wild-type Val is substituted by
FT Glu to reduce toxicity"
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992. 92MO-EP03016.
XX
XX 31-DEC-1991. 91IT-0MI3513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Pizsa M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX N-PSDB; AA051315.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX

```

```

CC The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of
CC this preferred detoxified mutain is not printed in the
CC specification but has been assembled from the full-length wild-type
CC sequence and the description given in the text. (N.B. Amino
CC acid numbering is based on the cholera toxin A subunit sequence).
XX
SQ Sequence 236 AA:
Query Match 66.7%; Score 10; DB 14; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SLRSAHLAQ 10
DB 64 SLRSAHLAQ 73
RESULT 27
AAR38732
ID AAR38732 standard; Protein; 236 AA.
AC AAR38732:
XX
XX 08-DEC-1993 (first entry)
DE "Tyr-53" E.coli heat labile toxin subunit A.
XX
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KM ADP-ribosyltransferase activity.
XX
XX Escherichia coli.
OS
FH Key Location/Qualifiers
FT Misc-difference 52 /note="corresponds to position 53 in cholera toxin
FT A subunit; wild-type Val is substituted by
FT Tyr to reduce toxicity"
XX
XX WO9313202-A.
XX
XX 08-JUL-1993.
XX
XX 30-DEC-1992. 92MO-EP03016.
XX
XX 31-DEC-1991. 91IT-0MI3513.
XX
XX (BIOC-) BIOCINE SCLAVO SPA.
XX
XX Domenighini M, Hol W, Pizsa M, Rappuoli R;
XX
XX WPI; 1993-227320/28.
XX
XX N-PSDB; AA051316.
XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin
PT - useful as vaccines against infection by Vibrio cholerae and
PT enterotoxin producing Escherichia coli
XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX
XX The wild-type sequence coding for the A subunit of the heat labile
CC toxin (LT-A) of a strain of E.coli known to affect humans (Yamamoto
CC et al, J.Biol. Chem., 259, 5037-5044 - see AA042768) was subjected to
CC site-directed mutagenesis. Certain mutations were found to reduce
CC toxicity (see AAR38730-32 and AAR44016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of

```

CC this preferred detoxified mutlein is not printed in the
 CC specification but has been assembled from the full-length wild-type
 CC sequence and the description given in the text. (N.B. Amino
 CC acid numbering is based on the cholera toxin A subunit sequence).

XX Sequence 236 AA;

Query Match 66.7%; Score 10; DB 14; Length 236;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQ 10
 |||||
 Db 64 SLRSAHLAQ 73

RESULT 28

AAM67772
 ID AAM67772 standard; Protein; 237 AA.

AC AAM67772;

DT 11-MAY-1999 (first entry)

DE E. coli heat labile toxin.

KW A subunit: heat labile toxin; ADP-ribosylation; mutant; detoxification;
 KW parenteral adjuvant; antigen; antigen; immunisation; humoral response;
 KW cell-mediated immune response; virus; bacterium; parasite; fungus;
 KW tumour; allergen; pathogen; AIDS; autoimmune disease; cancer; antibody;
 KW systemic lupus erythematosus; Alzheimer's disease; diagnosis.

OS Escherichia coli.

PN W09642375-A1.

PD 01-OCT-1998.

PF 19-MAR-1998; 98MO-US05454.

PR 18-MAR-1998; 98US-0044696.

PR 21-MAR-1997; 97US-0041227.

PA (CHIR) CHIRON CORP.

PI Barchfeld G, Del Giudice G, Rappuoli R;

DR WPI: 1999-070064/06.

DR N-PSDB: AAV81595.

PT Detoxified mutants of bacterial ADP-ribosylating toxins as
 PT parenteral adjuvants - useful to enhance humoral and cell-mediated
 PT immune responses in vertebrates when administered with selected
 PT antigen e.g. in disease treatment

PS Disclosure: Fig 1A-B; 51pp; English.

XX This sequence corresponds to the amino acid sequence of the A subunit of
 CC the E. coli heat labile toxin, an example of a bacterial ADP-ribosylating
 CC toxin. A mutant detoxified form of this protein is used in a parenteral
 CC adjuvant composition, which comprises the detoxified protein, at least
 CC one selected antigen and optionally a pharmaceutically acceptable
 CC (optionally topical) vehicle. The adjuvant composition can be
 CC administered parenterally in conjunction with at least one antigen in
 CC methods to immunise vertebrate subjects. The adjuvant has the ability
 CC to enhance the humoral and cell-mediated immune responses elicited by
 CC the antigen (e.g. by making the antigen more strongly immunogenic or
 CC necessitating fewer/lower antigen doses). It can be administered
 CC prior/subsequent to the antigen, and is preferably administered within
 CC a short space of time to the same site; it can also be administered in
 CC isolation from antigens as a boost following systemic or mucosal antigen
 CC administration. Most preferably, the adjuvant is co-administered with
 CC the antigen in the compositions and a pharmaceutically acceptable

CC carrier. The antigen may be derived from viruses, bacteria, parasites
 CC and fungi or may be tumour antigens, self-antigens and allergens. The
 CC compositions are therefore useful in the treatment and prevention of
 CC e.g. viral diseases, allergic manifestations, diseases caused by
 CC pathogens (e.g. bacteria or parasites), AIDS, autoimmune diseases
 CC (e.g. Systemic Lupus Erythematosus), Alzheimer's disease and cancers.
 CC The adjuvant can also be used to prepare antibodies against selected
 CC antigen(s), useful e.g. for diagnostic purposes or for antigen
 CC purification.

XX Sequence 237 AA;

Query Match 66.7%; Score 10; DB 20; Length 237;
 Best Local Similarity 100.0%; Pred. No. 0.0092;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQ 10
 |||||
 Db 65 SLRSAHLAQ 74

RESULT 29

AAU14105
 ID AAU14105 standard; peptide; 254 AA.

AC AAU14105;

DT 21-NOV-2001 (first entry)

DE Peptide sequence from Escherichia coli heat labile enterotoxin A.

KW Anti-retroviral; DP178-1like; DP107-1like; heat labile enterotoxin A;
 KW antitumorigenic; antiviral; HIV transmission.

OS Escherichia coli.

PN W0200151673-A2.

PD 19-JUL-2001.

PF 05-JUL-2000; 2000MO-US35727.

PR 09-JUL-1999; 99US-0350841.

PA (TRIM-) TRIMERIS INC.

PI Jeffs P, Lackey JW, Erickson JB, Lawless MK, Murtka G;

DR WPI: 2001-442157/47.

PT Identifying a compound that inhibits the formation of or disrupts a
 PT DP107/DP178 complex, especially compounds with antitumorigenic, antiviral
 PT or intracellular modulatory activity, by detecting the formation of a
 PT DP107/DP178 complex -

PS Disclosure: Fig 43; 259pp; English.

XX The present invention relates to peptides which exhibit anti-retroviral
 CC activity. The peptides of the invention (AAU12539-AAU14009) comprise
 CC DP178-1like and DP107-1like peptides. The DP178 peptide corresponds
 CC to amino acids 639-673 of the transmembrane protein gp41 from human
 CC immunodeficiency virus 1 (HIV-1) isolate LAI. The DP107 peptide
 CC corresponds to amino acids 558-595 of gp41 from HIV-1LAI. The invention
 CC also relates to a method of identifying compounds that inhibit the
 CC formation of or disrupts a DP107/DP178 complex. The method comprises
 CC detecting the formation of a DP107/DP178 complex, both in the presence
 CC or absence of a test compound, in a reaction mixture containing DP107
 CC and DP178 peptides. The method is useful for identifying compounds,
 CC including small molecule compounds, which may themselves exhibit
 CC antitumorigenic, antiviral or intracellular modulatory activity. The
 CC DP178-1like/DP107-1like peptides are useful to inhibit human and non-human
 CC retroviral, particularly HIV, transmission to uninfected cells. The
 CC present sequence represents a peptide sequence from Escherichia coli

Wed Jan 29 10:46:28 2003

CC heat labile enterotoxin A.
XX
SQ Sequence 254 AA:

Query Match 66.7%; Score 10; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.0097;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLRSAHLAQ 10
|||
DB 82 SLRSAHLAQ 91

Search completed: January 28, 2003, 10:32:17
Job time : 127 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 ; Search time 24.6667 Seconds
(without alignments)
58.460 Million cell updates/sec

Title: SEQ1-65TO79
Perfect score: 15
Sequence: 1 SLRSAHLAQSIILSG 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	15	100.0	258	1	OLECA heat-labile entero

ALIGNMENTS

RESULT 1
OLECA
heat-labile enterotoxin A precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 30-Apr-1981 #sequence_revision 17-Oct-1997 #extl_change 18-Jun-1999
C:Accession: I55231; A01817; A26946
R:Yamamoto, T.; Tamura, T.; Yokota, T.
J. Biol. Chem. 259, 5037-5044, 1984
A:Title: Primary structure of heat-labile enterotoxin produced by Escherichia coli patho
A:Reference number: I55231; MUID:84185610; PMID:6325417
A:Accession: I55231
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-258 <RES>
A:Cross-references: GB:K01995; NID:g148027; PIDN:AAA24685.1; PID:g148028
R:Spicer, E.K.; Noble, J.A.
J. Biol. Chem. 257, 5716-5721, 1982
A:Title: Escherichia coli heat-labile enterotoxin. Nucleotide sequence of the A subunit
A:Reference number: A01817; MUID:82167425; PMID:6279611
A:Accession: A01817
A:Molecule type: DNA
A:Residues: 1-21, 'R', 23-36, 'FNS', 40-44, 46-92, 'Y', 94-99, 'LTIY', 105-107, 111-118, 'IS', 121-
A:Cross-references: EMBL:V00275; NID:g41339; PIDN:CAA23532.1; PID:g41340

A>Note: the authors translated the codon TAT for residue 93 as Ser
R:Yamamoto, T.; Gojobori, T.; Yokota, T.
J. Bacteriol. 169, 1352-1357, 1987
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherich
A:Reference number: A26946; MUID:87137303; PMID:3546273
A:Accession: A26946
A:Molecule type: DNA
A:Residues: 1-21, 'R', 23-206, 'N', 208-230, 'E', 232-255, 'D', 257-258 <YAM>
A:Cross-references: EMBL:M15363
C:Comment: The heat-labile enterotoxin molecule contains one A chain and five or six
C:Gene: elta
C:Superfamily: heat-labile enterotoxin chain A
C:Keywords: enterotoxin
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-258/Product: heat-labile enterotoxin chain A #status predicted <MAT>

Query Match 100.0%; Score 15; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 9.6e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLRSAHLAQSIILSG 15
|||||
Db 83 SLRSAHLAQSIILSG 97

Search completed: January 28, 2003, 10:33:45
Job time : 24.6667 secs

Wed Jan 29 10:46:30 2003

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 ; Search time 11.6667 Seconds

(without alignments)
53.327 Million cell updates/sec

Title: SEQ1-65N079

Perfect score: 15

Sequence: 1 SLRSAMLAGOSTLSG 15

Scoring table: OLIGO

Searched: Gapop 60.0 , Gapext 60.0

Word size : 8

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	258	1 ELAH_ECOLI	P43530 escherichia
2	15	100.0	258	1 ELAP_ECOLI	P06717 escherichia

ALIGNMENTS

RESULT 1
ELAH_ECOLI
ID ELAH_ECOLI STANDARD; PRT; 258 AA.
AC P43530;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heat-labile enterotoxin A chain precursor (LT-A, human) (LTH-A).
GN ELTA OR LTPA OR TOXA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=078:H11 / H10407;
RX MEDLINE=84185610; PubMed=6325417;
RA Yamamoto T., Tamura T., Yokota T.;
RT "Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans";
RL J. Biol. Chem. 259:5037-5044(1984).
RN [2]
RP REVISION TO 207.
RC STRAIN=078:H11 / H10407;
RX MEDLINE=87137303; PubMed=3546273;
RA Yamamoto T., Gojobori T., Yokota T.;

RT "Evolutionary origin of pathogenic determinants in enterotoxigenic
RT Escherichia coli and Vibrio cholerae O1.";
RL J. Bacteriol. 169:1352-1357(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=078:H11 / H10407;
RX MEDLINE=93252225; PubMed=8486242;
RA Inoue T., Tsuji T., Koto M., Imanura S., Miyama A.;
RT "Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic Escherichia coli is identical to that of human strain H 10407.";
RL FEBS Microbiol. Lett. 108:157-161(1993).
RN [4]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=95349400; PubMed=7623669;
RA Domenighini M., Pizzo M., Jobling M.G., Holmes R.K., Rappuoli R.;
RT "Identification of errors among database sequence entries and comparison of correct amino acid sequences for the heat-labile enterotoxins of Escherichia coli and Vibrio cholerae.";
RL Mol. Microbiol. 15:1165-1167(1995).
CC -I- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -I- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; K01995; AAA24685.1; -;
CC DR EMBL; S60731; AAC60440.1; -;
CC DR HSSP; P06717; LTG.
CC DR InterPro; IPR001144; Enterotoxin_A.
CC DR Pfam; PF01375; Enterotoxin_A; 1.
CC DR PRINTS; PR00771; ENTEROTOXIN.
CC KM Enterotoxin; signal..
CC FT SIGNAL 1 18 BY SIMILARITY.
CC FT CHAIN 19 258 HEAT-LABILE ENTEROTOXIN A CHAIN.
CC FT DISULFID 205 217 BY SIMILARITY.
CC FT ACT_SITE 128 128 BY SIMILARITY.
CC FT CONFLICT 207 207 N -> D (IN REF. 1 AND 3).
CC SQ SEQUENCE 258 AA; 29872 MW; 2BB153C777FD78B7 CRC64;
Query Match 100.0%; Score 15; DB 1; Length 258;
Best local Similarity 100.0%; Pred. No. 6.4e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SLRSAMLAGOSTLSG 15
DB 83 SLRSAMLAGOSTLSG 97
RESULT 2
ELAP_ECOLI
ID ELAP_ECOLI STANDARD; PRT; 258 AA.
AC P06717; P01554;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Heat-labile enterotoxin A chain precursor (LT-A, porcine) (LTP-A).
GN ELTA OR LTPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate PC86, and Isolate P307;
RX MEDLINE=87137303; PubMed=3546273;
RA Yamamoto T., Gojobori T., Yokota T.;

```

RT "Evolutionary origin of pathogenic determinants in enterotoxigenic
RT Escherichia coli and Vibrio cholerae O1.";
RL J. Bacteriol. 169:1352-1357(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Isolate P307;
RA Dykes C.W., Halliday I.J., Hobden A.N., Read M.J., Harford S.;
RT "A comparison of the nucleotide sequence of the A subunit of heat-
RT labile enterotoxin and cholera toxin";
RL FEMS Microbiol. Lett. 26:171-174(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Isolate P307;
RA MEDLINE=82167425; PubMed=6279611;
RT Spicer E.K., Noble J.A.;
RT "Escherichia coli heat-labile enterotoxin. Nucleotide sequence of the
RT A subunit gene.";
RL J. Biol. Chem. 257:5716-5721(1982).
RN [4]
RP SEQUENCE OF 19-258 FROM N.A.
RC STRAIN-Isolate P307;
RA MEDLINE=91093102; PubMed=2266142;
RT Tsuji T., Inoue T., Miyama A., Okamoto K., Honda T., Miwatani T.;
RT "A single amino acid substitution in the A subunit of Escherichia
RT coli enterotoxin results in a loss of its toxic activity.";
RL J. Biol. Chem. 265:22520-22525(1990).
RN [5]
RP SEQUENCE OF 1-40 FROM N.A.
RA Trachman J.D., Maas W.K.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RA MEDLINE=93240541; PubMed=8478941;
RT Sikma T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;
RT "Refined structure of Escherichia coli heat-labile enterotoxin, a
RT close relative of cholera toxin.";
RL J. Mol. Biol. 230:890-918(1993).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RA MEDLINE=91238966; PubMed=2034287;
RT Sikma T.K., Pronk S.E., Kalk K.H., Wartna E.S., van Zanten B.A.M.,
RT Witvold B., Hol W.G.J.;
RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin
RT from E. coli.";
RL Nature 351:371-377(1991).
RN [8]
RP DISCUSSION OF SEQUENCE.
RA MEDLINE=93349400; PubMed=7623669;
RT Domenighini M., Pizzi M., Jobling M.G., Holmes R.K., Rappunoli R.;
RT "Identification of errors among database sequence entries and
RT comparison of correct amino acid sequences for the heat-labile
RT enterotoxins of Escherichia coli and Vibrio cholerae.";
RL Mol. Microbiol. 15:1165-1167(1995).
RN [9]
RP FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
RP THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
RN [10]
RP SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
RN [11]
RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M15361; AAA24791.1; -
CC EMBL: M15362; AAA24793.1; -
CC EMBL: M35581; AAA98202.1; -
CC EMBL: Y00275; CAA23532.1; -
CC EMBL: M57244; AAB59161.1; -
CC EMBL: M61015; AAA24335.1; -
CC EMBL: A04913; CAA00402.1; -
CC PIR: A26946; QLECEA.

```

```

DR PIR: A01817; QLECEA.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTA; 31-JUL-94.
DR PDB: 1LTA; 15-SEP-95.
DR PDB: 1LTA; 17-AUG-96.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTA; 07-JUL-97.
DR PDB: 1LTA; 16-JUN-97.
DR PDB: 1LTA; 20-APR-95.
DR InterPro: IPR001144; Enterotoxin_A.
DR Pfam: PF01375; Enterotoxin_A.1.
DR PRINTS: PR00771; ENTEROTOXINA.
KW Enterotoxin; Signal; 3D-structure.
FT SIGNAL 1 18
FT CHAIN 19 258
FT DISULFID 205 217
FT ACT_SITE 130 130
FT VARIANT 130 130
FT CONFLICT 37 39
FT CONFLICT 45 45
FT CONFLICT 93 93
FT CONFLICT 100 110
FT CONFLICT 119 120
FT CONFLICT 159 159
FT CONFLICT 207 207
FT STRAND 23 27
FT HELIX 31 37
FT TURN 36 38
FT STRAND 39 40
FT TURN 43 44
FT TURN 48 49
FT HELIX 59 64
FT TURN 75 76
FT STRAND 77 81
FT HELIX 84 95
FT STRAND 100 106
FT TURN 110 111
FT STRAND 112 114
FT HELIX 115 122
FT HELIX 126 128
FT STRAND 130 134
FT TURN 135 135
FT STRAND 137 138
FT STRAND 142 149
FT STRAND 150 151
FT STRAND 152 153
FT STRAND 158 159
FT TURN 161 162
FT HELIX 165 168
FT TURN 169 170
FT STRAND 174 174
FT HELIX 176 178
FT TURN 179 179
FT HELIX 180 182
FT TURN 187 188
FT HELIX 190 193
FT TURN 195 196
FT HELIX 197 200
FT TURN 203 204
FT HELIX 215 240
FT TURN 241 241
FT HELIX 242 244
FT HELIX 250 253
SO SEQUENCE 258 AA; 29902 MW; 2F0786442619F81F CRC64;
Query Match 100.0%; Score 15; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 6.4e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SLRSHLAGSISLG 15
|||||

```

Db 83 SLRSHLAGOSTISG 97

Search completed: January 28, 2003, 10:25:42
Job time : 12.6667 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 ; Search time 53.3333 Seconds
(without alignments)
57.951 Million cell updates/sec

Title: SEQ1-65TO79
Perfect score: 15
Sequence: 1 SLRSAHLAQSILSG 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP_mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP_rvirus:*
16: SP_bacteriaph:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	258	2 066280	066280 escherichia

ALIGNMENTS

RESULT 1
ID 066280 PRELIMINARY; PRT; 258 AA.
AC 066280;
DT 01-AUG-1998 (TRMBLrel. 07, Created)
DT 01-AUG-1998 (TRMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TRMBLrel. 17, Last annotation update)
DE Heat-labile enterotoxin A subunit.
GN LTH A SUBUNIT.
OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1032;
RA Komase K.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-1032;
RX MEDLINE-95091056; PubMed-7998417;
RA Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,
RA Hattori N., Watanabe K., Suzuki Y., Nagamine T., Aizawa C., Oya A.,
RA Kurata T.;
RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with
a trace amount of the holotoxin as an adjuvant for nasal influenza
vaccine.";
RT Vaccine. 12:1083-1089(1994).
RL Vaccine. 12:1083-1089(1994).
DR EMBL: AB011677; BAA25725.1; -.
DR HSSP: P06717; 1LTG.
DR InterPro: IPR001144; Enterotoxin_A.
DR InterPro: IPR000886; ER_Target.
DR Pfam: PF01375; Enterotoxin_A; 1.
DR PRINTS: PR00771; ENTEROTOXINA.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
SQ SEQUENCE 258 AA; 29931 MW; 2BB15D27740EB788 CRC64;

Query Match 100.0%; Score 15; DB 2; Length 258;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLRSAHLAQSILSG 15
Db 83 SLRSAHLAQSILSG 97

Search completed: January 28, 2003, 10:36:40
Job time : 53.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 23 Seconds
(without alignments)
19.189 Million cell updates/sec

Title: SEQ1-65to79
Perfect score: 15
Sequence: 1 SLRSAHLAQSIISG 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 8

Total number of hits satisfying chosen parameters: 11

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	240	US-08-823-120-3	Sequence 3, Appli
2	10	66.7	236	US-08-823-120-6	Sequence 6, Appli
3	10	66.7	234	US-08-486-099-114	Sequence 114, App
4	10	66.7	254	US-08-360-107A-114	Sequence 124, App
5	10	66.7	254	US-08-484-223B-114	Sequence 114, App
6	10	66.7	254	US-08-919-597-114	Sequence 114, App
7	10	66.7	254	US-08-475-668A-114	Sequence 114, App
8	10	66.7	254	US-08-485-551A-114	Sequence 114, App
9	10	66.7	254	US-08-471-913A-114	Sequence 114, App
10	10	66.7	254	US-08-485-264A-114	Sequence 114, App
11	10	66.7	254	US-08-474-349A-114	Sequence 114, App

ALIGNMENTS

RESULT 1
US-08-823-120-3
Sequence 3, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
APPLICANT: Pizsa, Mariagrazia
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
Cholera Toxin and of the Toxin Lt, Their Preparation and
TITLE OF INVENTION: Their Use for the Preparation of Vaccines

NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chilton Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-3
Query Match 100.0%; Score 15; DB 4; Length 240;
Best local Similarity 100.0%; Pred. No. 3 6e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SLRSAHLAQSIISG 15
Db 65 SLRSAHLAQSIISG 79
RESULT 2
US-08-823-120-6
Sequence 6, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
APPLICANT: Pizsa, Mariagrazia
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
Cholera Toxin and of the Toxin Lt, Their Preparation and
TITLE OF INVENTION: Their Use for the Preparation of Vaccines
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chilton Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:


```
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-484-223B-114

Query Match      66.7%; Score 10; DB 3; Length 254:
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SLRSAPHLAQ 10
        |||
        82 SLRSAPHLAQ 91

RESULT 6
US-08-919-597-114
; Sequence 114, Application US/08919597
; Patent No. 6054265
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
```

```
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,896
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-919-597-114

Query Match      66.7%; Score 10; DB 3; Length 254:
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SLRSAPHLAQ 10
        |||
        82 SLRSAPHLAQ 91

RESULT 7
US-08-475-668A-114
; Sequence 114, Application US/08475668A
; Patent No. 6060065
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,668A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
```

REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-475-668A-114

Query Match
Best Local Similarity 100.0%; Score 10; DB 3; Length 254;
Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIRSASLAGQ 10
|||||
Db 82 SIRSASLAGQ 91

RESULT 8
US-08-485-551A-114
Sequence 114, Application US/08485551A
Patent No. 6068973
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Mathews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Peteway, Stephen R.
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,551A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-023
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-551A-114

Query Match
Best Local Similarity 100.0%; Score 10; DB 3; Length 254;
Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIRSASLAGQ 10
|||||
Db 82 SIRSASLAGQ 91

RESULT 9
US-08-471-913A-114
Sequence 114, Application US/08471913A
Patent No. 6093794
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Mathews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Peteway, Stephen R.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 214
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,913A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-471-913A-114

Query Match
Best Local Similarity 100.0%; Score 10; DB 3; Length 254;
Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIRSASLAGQ 10
|||||
Db 82 SIRSASLAGQ 91

RESULT 10
US-08-485-264A-114
Sequence 114, Application US/08485264A
Patent No. 6228983
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.

APPLICANT: Mathews, Thomas J.
 APPLICANT: Wild, Carl T.
 APPLICANT: Barney, Shawn O.
 APPLICANT: Lambert, Dennis M.
 APPLICANT: Pellew, Stephen R.
 APPLICANT: Langlois, Alphonse J.
 TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
 TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
 TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
 NUMBER OF SEQUENCES: 232
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 115 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/485,264A
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7872-021
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-9741/8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 114:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 254 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-485-264A-114

QY 1 SURSAHLAQ 10
111111111
Db 82 SURSAHLAQ 91

RESULT 11
US-08-474-349A-114
Sequence 114, Application US/08474349A
Patent No. 6333395

GENERAL INFORMATION:

APPLICANT: Bolognesi, Daniel P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Peteway, Stephen R.
APPLICANT: Langlois, Alphonse J.

TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
NUMBER OF SEQUENCES: 517

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.300
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,349A
FILING DATE: 07-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ. ID NO.: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-474-349A-114

Query Match	Score 10;	DB 4;	Length 25;
Best Local Similarity	100.0%;	Pred. No. 0.004;	
Matches 10;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps 0;			
QY	1	SURSAHLAQ	10
Db	82	SURSAHLAQ	91

Search completed: January 28, 2003, 10:38:03
Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:25:45 ; Search time 13.3333 Seconds
(without alignments)
22.701 Million cell updates/sec

Title: SEQ1-65TO79

Perfect score: 15
Sequence: 1 SLRSAHLAQGSILSG 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 122226 seqs, 20178551 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications-AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	240	10	US-09-819-917-7 Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-819-917-7
; Sequence 7, Application US/09819917
; Patent No. US20020044939A1
; GENERAL INFORMATION:
; APPLICANT: Pizze, Mariagrazia
; APPLICANT: Fontana, Maria Rita
; APPLICANT: Giannelli, Valentina
; APPLICANT: Rappuoli, Rina
; TITLE OF INVENTION: Immunogenic Detoxified Mutants Of Cholera Toxin
; FILE REFERENCE: CHIR0312
; CURRENT APPLICATION NUMBER: US/09/819,917
; CURRENT FILING DATE: 2001-03-28

; PRIOR APPLICATION NUMBER: 08/981,208
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: 9513371.6
; PRIOR FILING DATE: 1995-06-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 240
; TYPE: PRT
; ORGANISM: E. coli
US-09-819-917-7

Query Match 100.0%; Score 15; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLRSAHLAQGSILSG 15
|||||
Db 65 SLRSAHLAQGSILSG 79

Search completed: January 28, 2003, 10:38:58
Job time : 13.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 ; Search time 127 Seconds
(without alignments)
15.738 Million cell updates/sec

Title: SEQ3-65TO79
Perfect score: 15
Sequence: 1 STFEQVPNNKEFGV 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

A_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
No matches found				

Search completed: January 28, 2003, 10:32:17
Job time : 127 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 23 Seconds
(without alignments)
19.189 Million cell updates/sec

Title: SEQ3-65T079
Perfect score: 15
Sequence: 1 STFEQVPNNKEFKGV 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/DackR11est1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	53.3	241	4	US-08-823-120-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-823-120-1
; Sequence 1, Application US/08823120
; Patent No. 6149919
; GENERAL INFORMATION:
; APPLICANT: Domenighini, Mario
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; TITLE OF INVENTION: Immunogenic Detoxified Mutants of
; TITLE OF INVENTION: Cholera Toxin and of the Toxin Lt, Their Preparation and
; TITLE OF INVENTION: Their Use for the Preparation of Vaccines
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 53.3% Score 8: DB 4: Length 241;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 8: Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 6 VPNNKEFK 13
|||
DB 186 VPNNKEFK 193

Search completed: January 28, 2003, 10:38:03
Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:25:45 ; Search time 13.3333 Seconds
(without alignments)
22.701 Million cell updates/sec

Title: SEQ3-65TO79

Perfect score: 15
Sequence: 1 STFEQVPNNKEFKGV 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 122226 seqs, 20178551 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	53.3	241	10	US-09-819-917-5 Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-09-819-917-5
; Sequence 5, Application US/09819917
; Patent No. US20020044939A1
; GENERAL INFORMATION:
; APPLICANT: Piza, Mariagrazia
; APPLICANT: Fontana, Maria Rita
; APPLICANT: Giannelli, Valentina
; APPLICANT: Rappunli, Rina
; TITLE OF INVENTION: Immunogenic Detoxified Mutants Of Cholera Toxin
; FILE REFERENCE: CHIR0312
; CURRENT APPLICATION NUMBER: US/09/819,917
; CURRENT FILING DATE: 2001-03-28

; PRIOR APPLICATION NUMBER: 08/981,208
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: 9513371.6
; PRIOR FILING DATE: 1995-06-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 241
; TYPE: PRT
; ORGANISM: E. coli
US-09-819-917-5

Query Match 53.3%; Score 8; DB 10; Length 241;
Best local Similarity 100.0%; Pred. No. 0.039;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VPNNKEFK 13
|||
Db 186 VPNNKEFK 193

Search completed: January 28, 2003, 10:38:58
Job time : 13.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 24.6667 Seconds
(without alignments)
58.460 Million cell updates/sec

Title: SEQ3-65TO79
Perfect score: 15
Sequence: 1 STFEQVPNNKEFGV 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*
1: plr1:*
2: plr2:*
3: plr3:*
4: plr4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
1	9	60.0	259 2 A29831	heat-labile entero

ALIGNMENTS

RESULT 1
A29831
heat-labile enterotoxin IIA chain A precursor - Escherichia coli
N:Alternate names: LT-IIa
C:Species: Escherichia coli
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A29831
R:Pickett, C.L.; Weinstein, D.L.; Holmes, R.K.
J. Bacteriol. 169, 5180-5187, 1987
A:Title: Genetics of type IIA heat-labile enterotoxin of Escherichia coli: operon fusion
A:Reference number: A91849; MUID:88032841; PMID:2822667
A:Accession: A29831
A:Molecule type: DNA
A:Residues: 1-259 <PIC>
A:Cross-references: GB:M17894; MID:9146671; PIDN:AAA24093.1; PID:9146672
A:Note: the authors translated the codon TAT for residue 225 as Thr
C:Superfamily: heat-labile enterotoxin chain A
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-259/Product: heat-labile enterotoxin IIA chain A #status predicted <LTA>

Query Match 60.0%; Score 9; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 0.0056;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VPNNKEFG 14
|||
Db 204 VPNNKEFG 212

Search completed: January 28, 2003, 10:33:46
Job time : 25.6667 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 11.6667 seconds
(without alignments)
53.327 Million cell updates/sec

Title: SEQ3-65TO79

Perfect score: 15
Sequence: 1 STFEEVPMNKEFKG 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	60.0	259	1 E2AA_ECOLI	P13810 escherichia

ALIGNMENTS

```

RESULT 1
E2AA_ECOLI          STANDARD;          PRT;          259 AA.
AC P13810;
DT 01-JAN-1990 (rel. 13, Created)
DT 01-JAN-1990 (rel. 13, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Heat-labile enterotoxin IIA, A chain precursor (LT-IIA).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=8032841; PubMed=2822667;
RA Pickett C.L., Weinstein D.L., Holmes R.K.;
RT "Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
RT operon fusions, nucleotide sequence, and hybridization studies.";
RL J. Bacteriol. 169:5180-5187(1987).
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -!- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
  
```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

-----
DR EMBL; M17894; AAA24093.1; -.
DR PIR; A29831; A29831.
DR HSSP; P43528; 1TII.
DR InterPro; IPR001144; Enterotoxin_A.
DR Pfam; PF01375; Enterotoxin_A; 1.
DR PRINTS; PR00771; ENTEROTOXINA.
KW Enterotoxin; Signal.
FT SIGNAL 1 18
FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
FT DISULFID 203 215 BY SIMILARITY.
FT ACT_SITE 128 128 BY SIMILARITY.
SQ SEQUENCE 259 AA; 29242 MW; 996F311A32CABEAA CRC64;
  
```

Query Match 60.0%; Score 9; DB 1; Length 259;
Best local Similarity 100.0%; Pred. No. 0.0021;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6 VPNNKEFKG 14
   |||||
DB 204 VPNNKEFKG 212
  
```

Search completed: January 28, 2003, 10:25:42
Job time : 11.6667 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 53.3333 Seconds
(without alignments)
57.951 Million cell updates/sec

Title: SEQ3-65TO79
Perfect score: 15
Sequence: 1 STFEQVPNNKEFKGV 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match length DB ID Description

No matches found		

Search completed: January 28, 2003, 10:36:40
Job time : 53.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 127 Seconds
(without alignments)
15.738 Million cell updates/sec

Title: SEQ4-65TO79
Perfect score: 15
Sequence: 1 REFNSLPNNKASSDF 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

A_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

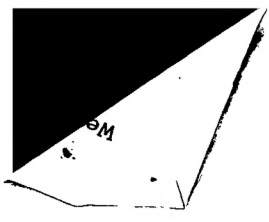
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description

No matches found			

Search completed: January 28, 2003, 10:32:17
Job time : 127 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 ; Search time 24.6667 seconds
(without alignments)
58.460 Million cell updates/sec

Title: SEQ4-65TO79
Perfect score: 15
Sequence: 1 REFNSLPNNKASDPT 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

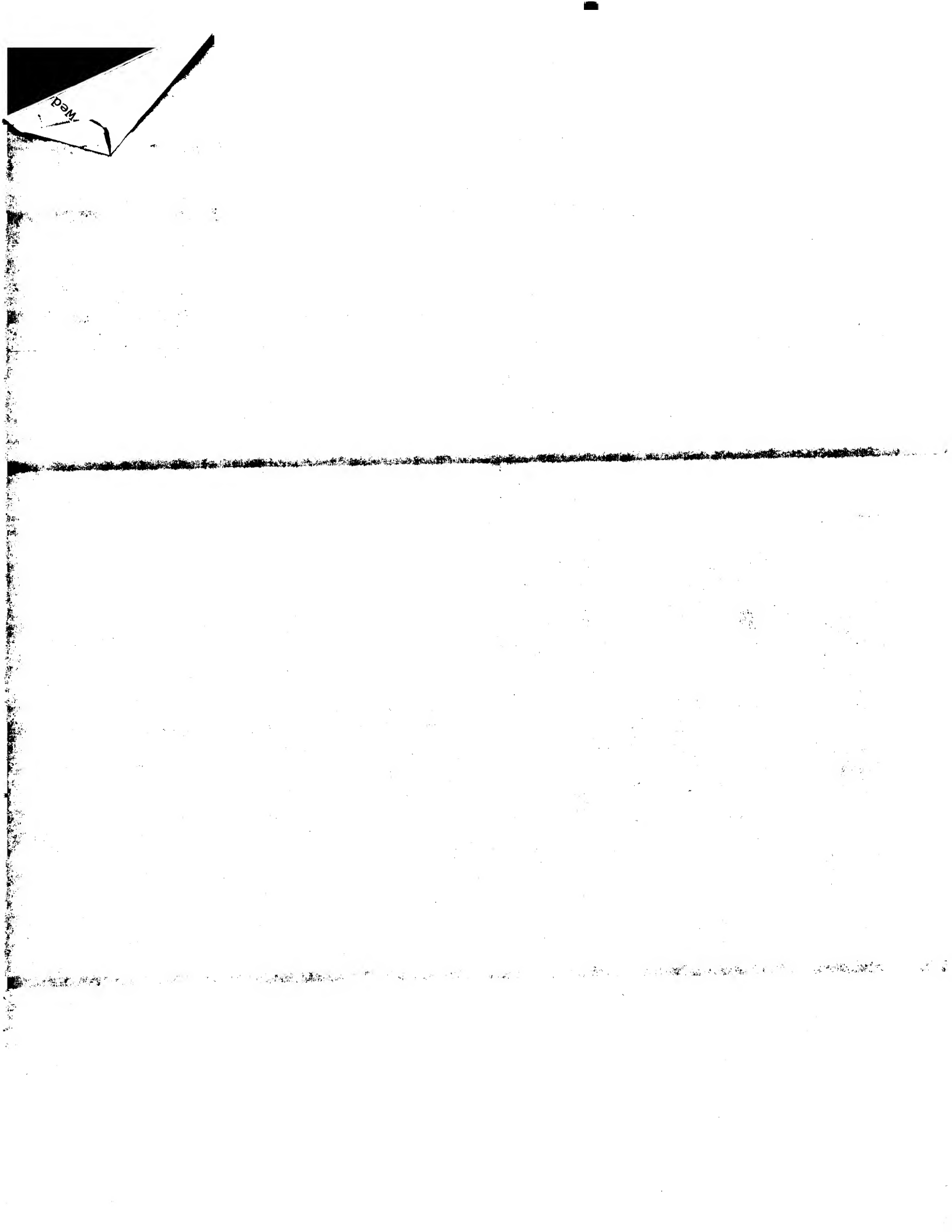
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB	ID	Description
---------------	----------------	-------	--------	----	----	-------------

No matches found

Search completed: January 28, 2003, 10:33:46
Job time : 24.6667 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 11.6667 seconds
(without alignments)
53.327 Million cell updates/sec

Title: SEQ4-655T079

Perfect score: 15
Sequence: 1 REFNSLPNNKASDPT 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	10	66.7	263	1 E2BA_ECOLI	P43528 escherichia

ALIGNMENTS

RESULT 1
E2BA_ECOLI STANDARD; PRT; 263 AA.
AC P43528;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heat-labile enterotoxin IIB, A chain precursor (LT-IIB).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 41;
RX MEDLINE=89359131; PubMed=2670900;
RA Pickett C.L., Twiddy E.M., Coker C., Holmes R.K.;
RT "Cloning, nucleotide sequence, and hybridization studies of the type
RT IIB heat-labile enterotoxin gene of Escherichia coli.";
RL J. Bacteriol. 171:4945-4952(1989).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).
RX MEDLINE=96399713; PubMed=8805549;
RA van den Akker F., Sarfaty S., Twiddy E.M., Connell T.D., Holmes R.K.,
RA Hol W.G.J.;
RT "Crystal structure of a new heat-labile enterotoxin, LT-IIB.";
RL Structure 4:665-678(1996).

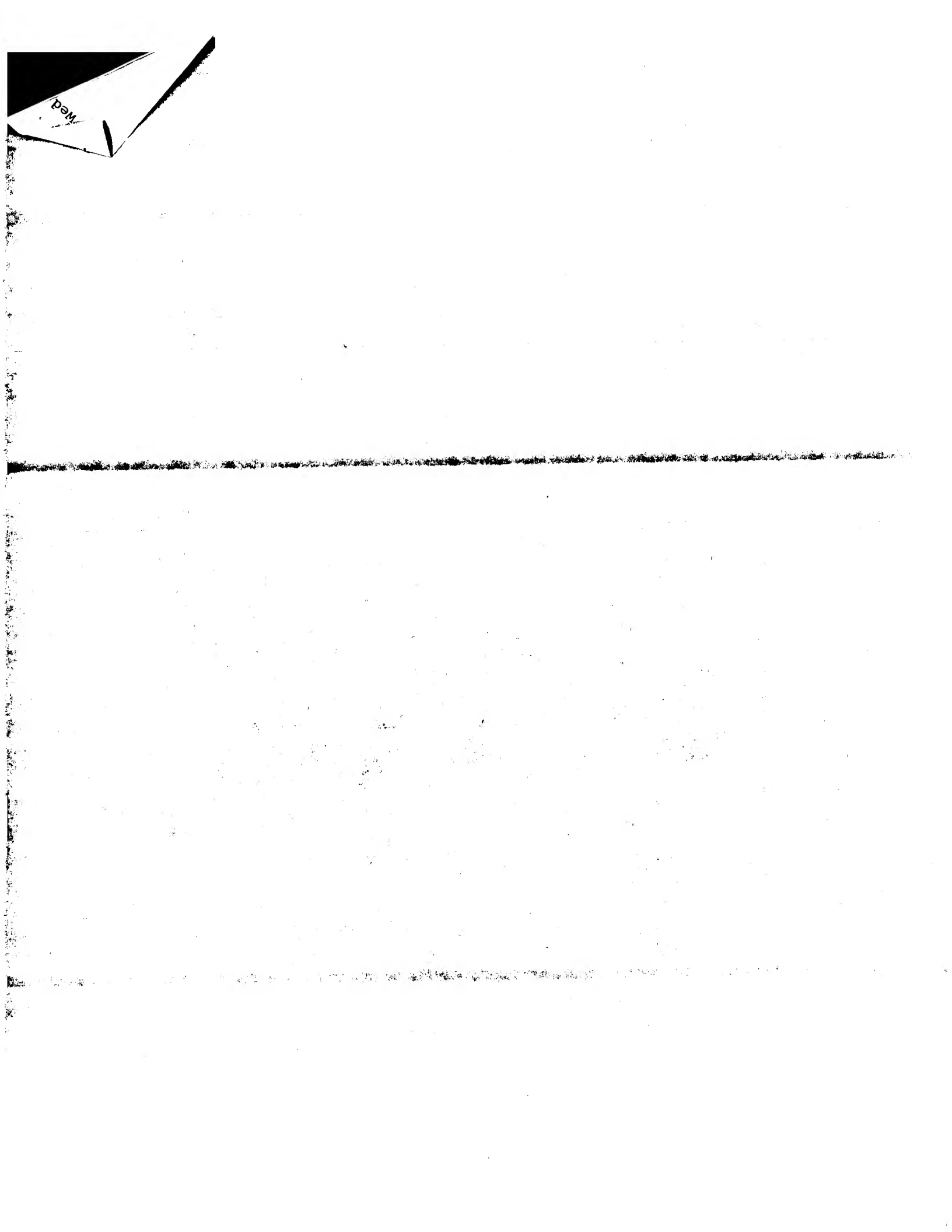
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -!- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; M28523; AA53285.1; .
DR PDB; 1RTI; 17-AUG-96.
DR InterPro: IPR001144; Enterotoxin_A.
DR Pfam: PF01375; Enterotoxin_A; 1.
DR PRINTS; PR00771; ENTEROTOXINA.
KW Enterotoxin; Signal; 3D-structure.
FT SIGNAL 1 263
FT CHAIN 21 263 HEAT-LABILE ENTEROTOXIN IIB, A CHAIN.
FT DISULFID 205 217
FT ACT_SITE 130 130
SQ SEQUENCE 263 AA; 29485 MW; A56E8A069427CEB9 CRC64;

Query Match 66.7%; Score 10; DB 1; Length 263;
Best local Similarity 100.0%; Pred. No. 0.00064;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 LPNNKASDPT 15
DB 206 LPNNKASDPT 215

Search completed: January 28, 2003, 10:25:42
Job time : 11.6667 secs



GenCore version 5.1.3
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 ; Search time 53.3333 Seconds
 (without alignments)
 57.951 Million cell updates/sec

Title: SEQ4-65TO79
 Perfect score: 15
 Sequence: 1 REFNSLPNNKASSDT 15

Scoring table: OLIGO
 Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

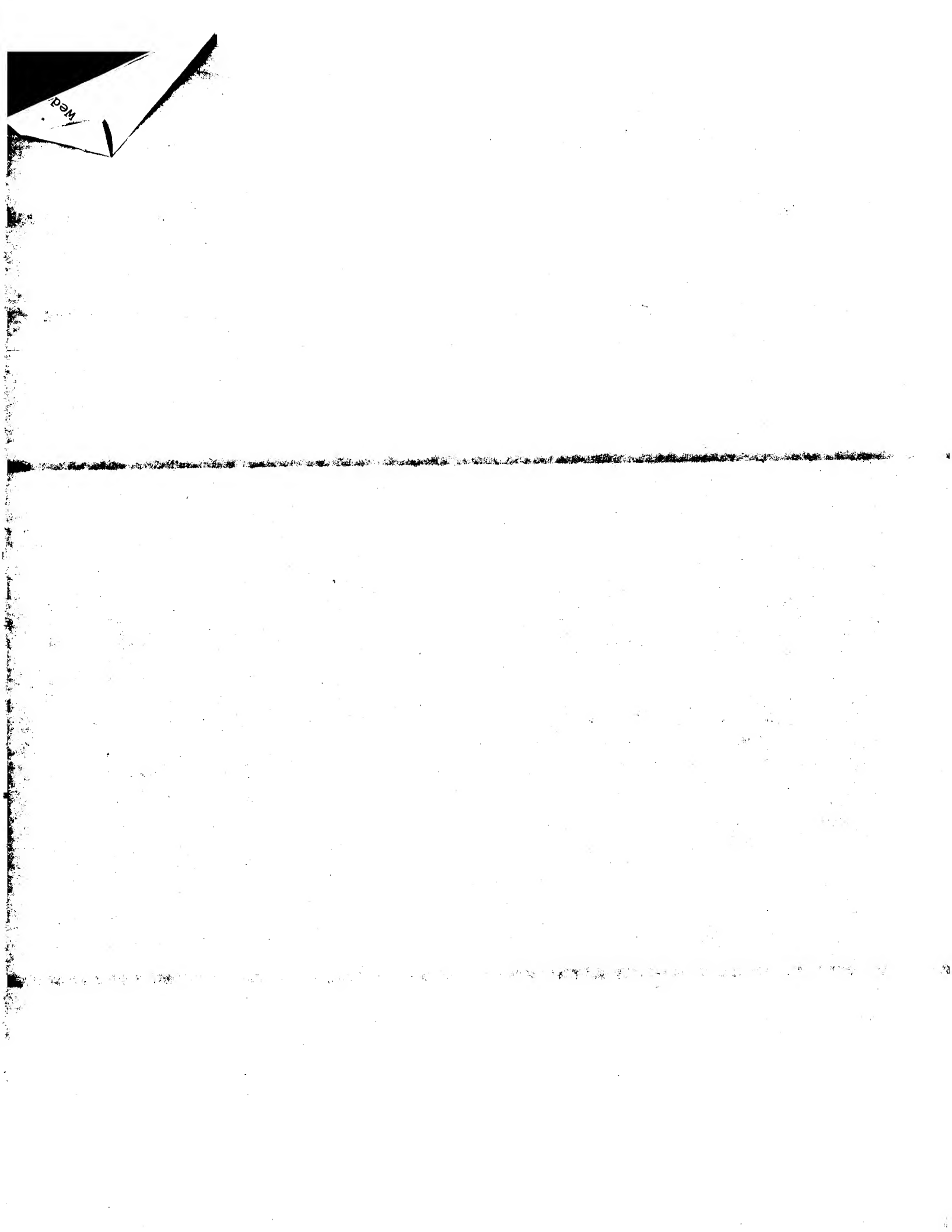
SPTREMBL_21:*
 1: sp_archaea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phage:*
 10: sp_plant:*
 11: sp_rodent:*
 12: sp_virus:*
 13: sp_vertebrate:*
 14: sp_unclassified:*
 15: sp_rvirus:*
 16: sp_bacteriaph:*
 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
No matches found				

Search completed: January 28, 2003, 10:36:40
 Job time : 53.3333 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 28, 2003, 10:24:52 : Search time 23 Seconds
(without alignments)
19.189 Million cell updates/sec

Title: SEQ4-65TO79
Perfect score: 15
Sequence: 1 REFNSLPNNKASSDF 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
- 4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
- 5: /cgn2_6/prodata/2/1aa/PCUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/1aa/Dackliles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				Description
No.	Score	Match	Length	ID	

No matches found					

Search completed: January 28, 2003, 10:38:03
Job time : 23 secs

Med

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 28, 2003, 10:25:45 : Search time 13.3333 Seconds
(without alignments)
22.701 Million cell updates/sec

Title: SEQ4-65TO79
Perfect score: 15
Sequence: 1 REFNSLPNNKASSDR 15

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 122226 seqs, 20178551 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description

No matches found

Search completed: January 28, 2003, 10:38:58
Job time : 13.3333 secs

